

The politics of plutonium disposal

The new US policy on plutonium disposition is the right one, and efforts to encourage similar action in Russia must be redoubled.

THE decision by the United States to pursue a 'dual track' policy for the disposal of excess plutonium from nuclear weapons is a difficult and necessary step on the long and rocky road to winding down the weapons stockpiles in the United States and Russia. But the ratcheting down of the world's vast stock of fissile material is a task at once more urgent and more complicated than is generally acknowledged.

Hazel O'Leary, the US energy secretary, announced last week that her department will prepare for two means of plutonium disposal simultaneously. It will work out how to mix the element with other, highly radioactive waste and fuse it into large glass or ceramic bricks for storage and eventual burial. At the same time, it will explore the processing of plutonium into mixed oxide (MOX) fuel for burning in civilian nuclear power plants, and subsequent disposition with other spent nuclear fuel.

Left to its own devices, the United States would probably have stuck to the first option. A recently opened vitrification plant at Savannah River, South Carolina, could do the job quite comfortably. The United States has a twenty-year-old policy of keeping the military and civil nuclear-fuel cycles separate from each other, which the MOX option will breach. This administration has no domestic incentive to make trouble with the environmental movement by modifying that policy and distributing MOX containing military plutonium to power plants around the country.

But Russia has an even bigger inventory of excess military plutonium than the United States, and that material is more likely to fall into dangerous hands. Both countries have more than 100 tonnes of military plutonium, of which half is surplus to current military needs. The excess is stored as dismantled weapons cores, or 'pits'. Each weighs about 5 kg and could be delivered to, say, central London in the glove compartment of a car.

But where the United States sees the excess plutonium as a dangerous liability, Russia (in common with France, Britain and Japan) views it as a potentially valuable energy source. Russia is not prepared to bury it in glass. Had the United States chosen to do so, Russia could have ignored the move and kept its excess plutonium in the warehouse. It could even have made a plausible case that the United States was technically equipped to retrieve the plutonium from glass, in time of need.

An independent commission of Russian and US scientists — established earlier this year at the suggestion of Boris Yeltsin, the Russian president — recently pointed out that global security will be best served if Russia and the United States work in tandem on this problem. That means that both countries should pursue both vitrification and MOX fuel burning, the commission said.

Fortunately, the US administration has accepted that advice, and shelved domestic considerations in favour of the larger picture. Its decision has attracted the expected flak from environmentalists. Respected critics of the plutonium economy, such as Paul Leventhal of the Nuclear Control Institute, have attacked it as an "endorsement of plutonium" by the United States, which will, by example, spread the worldwide use of plutonium as a fuel.

These criticisms are not to be taken lightly, but they ultimately represent a triumph of dogmatism over pragmatism. The United States still has 6,500 nuclear weapons deployed, and a larger number in reserve, and spends \$4 billion a year retaining the capacity

to develop still more. The idea that it can lead the world on this issue by moral example is, beyond the confines of a Washington seminar room, laughable. What matters are its actions — in this case, its promise to dispose of its plutonium stock expeditiously while making the compromises necessary to get Russia to do the same. With the US policy in place, the path is now open for the United States to reach agreement rapidly with Russia on the disposal of excess military plutonium, along the lines recommended by the Russian-US commission of scientists, which will report fully in the new year.

Implementation of the US scheme will be difficult enough (although Canada may take the MOX if US power plants cannot be bribed to accept it). Orchestrating the political will and technical means to dispose of weapons plutonium in Russia will present a far sterner challenge. Any solution will have to draw on extensive Western technical and financial assistance. The latter will require the approval of the US Congress, which, despite the best efforts of internationalists such as Senator Richard Lugar (Republican, Indiana), has become increasingly insular and curmudgeonly of late in its approach to aid for Russia.

O'Leary's decision contains — to the environmentalists' horror — some incentives to win congressional support, such as the possibility of a new MOX plant at Savannah River. Hopefully Strom Thurmond (Republican, South Carolina), chair of the Senate Armed Services Committee, has the vision to do the right thing by Russia, with or without such crude incentives.

But even an agreement to dispose of 50 tonnes of excess plutonium in both countries would be only one further step in reducing the world's plutonium inventory. There are, after all, the 50 tonnes still incorporated in the military stockpiles of each country. A START-3 agreement to reduce weapons levels further is awaiting ratification of START 2 by the Russian Duma.

There is, moreover, the growing global inventory of civil plutonium, which now totals around 1,000 tonnes, four-fifths of it in spent fuel and the rest as plutonium oxide or MOX. This material is far less bomb-ready than the pits piling up in Russia and the United States, but it still represents a substantial proliferation hazard. Unlike highly enriched uranium, which can be extracted only by isotopic separation, plutonium could be extracted by relatively simple chemical means. The isotopic make-up of civil plutonium is different from that of weapons plutonium, but it can still be used to make a bomb, as the United States proved in a test in 1962.

The United States has sought to discourage the separation and reprocessing of yet more plutonium at commercial plants in Britain, France and, soon, Japan. It should proceed with plutonium destruction in commercial reactors, while opposing the separation of new plutonium with even greater vigour than before. The nuclear genie will not be back in the bottle until reprocessing plants such as THORP in the United Kingdom are closed down — but that can only be regarded as a long-term objective.

In the meantime, Russia is the main proliferation issue. O'Leary has moved with some skill to neutralize parties in the United States hostile to foreign aid and oblivious to the importance of improved management of Russian nuclear material. The onus is now on the Congress to provide the modest sums of money needed to support plutonium disposition at home and abroad. □



Goldin wants more NASA biologists as Gore is briefed on space plans

Washington. Dan Goldin, administrator of the US National Aeronautics and Space Administration (NASA), said last week that the agency should significantly increase its research on astrobiology — the science of the origins of life — and employ more biologists.

Goldin was speaking after a meeting of leading space scientists with Al Gore, the US vice-president, which not only covered recent and possible future discoveries in space science but also addressed their potential religious and philosophical implications.

"We were working around the core theme of origins, including the origins of life," said one participant, Claude Canizares of the Massachusetts Institute of Technology, who chairs a National Academy of Sciences space studies board. He said the religious perspective was required because "this touches on the most profound philosophical questions".

The meeting had been organized to brief Gore on the US space science programme in preparation for a US 'space summit' in February. Among those invited to participate were John Minogue, a theologian and president of DePaul University in Chicago, Stephen Jay Gould, the evolutionary biologist, and Carl Sagan, the popular astronomer and author (although Sagan was ill and unable to attend).

The two-and-a-half hour meeting was described by officials as a wide-ranging

discussion, led by Gore, of the "philosophical underpinnings" of space science, which steered clear of specifics. But Goldin said that the meeting had taught him that the space agency's small programme in astrobiology should be expanded.

"It became clear to me that the astrobiology programme is desperately underfunded," Goldin said. "We had better look at our science on the origins of life, before we go building more spacecraft."

Warming to this theme, Goldin said that life scientists are underrepresented at NASA's research centres. "We have too many people in chemistry and physics," he said, adding that in future the agency "will be led by two disciplines — information science and biology".

NASA currently has a small programme in the new discipline of astrobiology at the Ames Research Centre in California, led by David Morrison. A spokesman at NASA headquarters in Washington was unable to provide a detailed breakdown of the agency's current research or staff by scientific discipline.

Jack Gibbons, science adviser to President Bill Clinton, says last week's meeting did not discuss programmes or budgets. He also plays down expectations that the summit will make decisions on specific programmes. "I'm not optimistic we will get that degree of specificity by February," he says. The summit will bring Clinton and Gore



together with congressional leaders to discuss an impending budget squeeze at NASA.

The summit was first proposed by Senator Barbara Mikulski (Democrat, Maryland), a NASA supporter, and agreed to by Clinton in August, when the agency's scientists claimed to have found evidence of life on meteorites that originated on Mars.

The group of 20 that met Gore last week also included David McKay of the NASA Johnson Space Center in Texas, lead author of the meteorite paper (*Science* 273, 924; 1996); Bill Moyers, a television commentator; Stuart Kauffman of the Santa Fe Institute, an expert on complexity theory; and Anneila Sargent of the California Institute of Technology, who chairs NASA's space science advisory board.

The presence of religious leaders indicates the significance now attached to this aspect of space exploration. "Many Americans believe in God," said Goldin. "Taxpayer funds are involved here and it is crucial that science is not operating in a vacuum."

But Gould is sceptical. "We should acknowledge that religion and science are different enterprises that do not intersect," he says. Science oversteps its boundaries, he suggests, "when it tries to talk about the fundamental nature of the Universe", and religion oversteps its own when it asserts that nothing is more than 10,000 years old.

Gore's unusually prominent role in the Clinton administration — his responsibilities include environmental policy and relations with Russia — is further enhanced by the expectation that he will mount a formidable bid for the presidency in 2000. His strong intellectual grasp of the issues made a deep impact on delegates at last week's meeting. "There wasn't a scientist in the room who was not impressed," says Sargent.

Colin Macilwain

NIH drops plan for multiyear grants

Washington. The US National Institutes of Health (NIH) has abandoned a plan to issue some young researchers with multiyear grants after lawyers at the Department of Health and Human Services said that it would breach departmental regulations.

The reversal means that the NIH will now have to spend all of its 1997 research funds next year — rather than holding some back to pay for research in future years — so boosting the number of grants it will make.

The biomedical research agency's plan had been to select about 250 young investigators and support them for two or three years. But it decided to drop the idea when lawyers advised that it would have to give multiyear grants to all the successful applicants for certain categories of grants or to none at all.

NIH had argued that the multiyear grants would guarantee recipients that they would continue to receive funding, and would prevent the agency from issuing too many new grants in 1997 — a year for which the agency has received a generous funding increase. But biomedical lobbyists attacked the idea on the grounds that it would allow NIH to avoid spending all of the money that Congress has given it to spend, so weakening its case for more money in subsequent years (see *Nature* 384, 203; 1996).

Wendy Baldwin, deputy director for extramural research at NIH, says that the policy change came about "abruptly" at a time when the agency's institutes had been working out how to issue the multiyear grants. A handful have already been awarded, and these will now be given as single-year grants. **C. M.**

Canada's HIV blood inquiry turns poisonous

Montreal. Lawyers representing Canada's Liberal federal government have launched a fierce attack on an inquiry set up by its Progressive Conservative predecessor into the country's system of blood distribution, just as the inquiry is entering its final stage.

And internal changes being introduced by the Canadian Red Cross in advance of the inquiry's report have been criticized for reducing the influence of physicians on the way the system is run.

The inquiry, headed by Mr Justice Horace Krever, is seeking to allocate blame for the infection of some 1,200 Canadians with the human immunodeficiency virus (HIV) and another 12,000 with hepatitis C through contaminated blood and blood products (see *Nature* 379, 663; 1996).

Together with the Red Cross — as well as various provinces, pharmaceutical companies and individuals — the government has already appealed to the courts to limit Krever's conclusions, and has denied him access to some documents he requested. It has also set up a task force to submit a plan by 15 February to remodel the federal blood system — pre-empting Krever's report, which is not due until the end of April.

In a new twist, the federal government has now accused Krever of looking for scapegoats rather than focusing on improvements to the blood distribution system, and has called his potential findings of misconduct against bureaucrats "repugnant" and "without foundation".

The charges are contained in the government's final submission to the inquiry. But

David Dingwall, the health minister, promptly dissociated himself, the government and the justice minister from its "inflammatory language" — which he blamed on lawyers in the justice department.

This was followed by an apology to Krever from the federal government lawyer, Donald Rennie, who admitted that the language of the government's submission could be interpreted as impugning the judge's reputation and that of his inquiry — although denying that this had been its intention.

The Canadian Red Cross, in its own final submission to the inquiry commission, says that individuals or institutions should not be blamed for what happened because those involved, "without exception, were doing the best they could". Yet in the same submission it argues that the federal Canadian Blood Agency was responsible for a delay in implementing HIV testing of the blood supply, for which the Red Cross has been blamed.

Without waiting for the recommendations of either the commission or the government task force, the Red Cross has already made far-reaching changes in its internal organization. But these have been controversial. Fundamental differences of opinion between physicians and bureaucrats have led to the departure of about a third of the agency's physicians, and others are known to be dissatisfied.

One of those who have left in protest is Bert Aye, former director of Red Cross national blood services. Aye says that he could not support the changes, either morally or professionally. "The Red Cross

has systematically removed all physicians who were in charge of blood centres — including the national director — and replaced them with lay managers," he says.

Aye argues that it is the physicians who continue to bear responsibility for maintaining the safety and quality of blood supplies. "The authority and role of the physicians in the new structure remain ambiguous, yet their responsibility and accountability remain well defined. What will happen if and when another crisis appears such as that of HIV?" he asks. "Who will take responsibility? A committee?"

The Red Cross's reorganization seems to imply that the blood contamination was the fault of doctors, says Aye, and that replacing them with laypeople should rectify the problem. But, he says, laymen do not have the technical expertise to make the system safer. In fact, as new and more complex biotechnological procedures are being adopted, he says, the need for closer involvement of physicians increases.

Aye fears that current plans for reshaping blood systems in Canada and other countries would not improve the safety of the blood supply or reduce the impact of future crises. "Most of the management reviews in Canada, the United States and the United Kingdom have been done either with the objective of containing costs or to improve regulatory compliance," he says. "These are limited objectives and, if continued, will mean that blood centres will no longer be able to attract the kind of professionals required."

David Spurgeon

'Ban tourists' call to save glacier source of the Ganges

New Delhi. The Gangotri glacier in the Himalayas, the source of the river Ganges, is threatened by tourist traffic, adventure sports and pilgrims visiting nearby shrines, say Indian glaciologists. They want a ban on human activities that they believe accelerate the process of weathering of the glacier, which has shrunk by 600 metres since 1935.

"Massive cracks have developed in its lower reaches," says Syed Iqbal Husnain, a professor in the department of environmental sciences at Jawaharlal Nehru University in New Delhi. Husnain led an expedition to the glacier recently and found that about 3 square kilometres of ice had collapsed.

Husnain runs the Himalayan glaciers project, funded by India's science ministry, and also heads the Asian chapter of the International Commission for Snow and Ice. He says his team found "glaring changes" in Gangotri since his last survey in 1991. "There are huge blocks of dead ice covered by debris completely detached from the main body and the snout has virtually collapsed." All Himalayan glaciers are

receding, but at Gangotri human intervention and environmental degradation are making the problem worse, he says.

A lush green valley at Topoban has become an attraction for tourists and trekkers, who must cross the glacier to reach it. The glacier is popular with package tours as it is easily approachable. "Each year half a million pilgrims visit a Hindu shrine 15 km down below," says D. N. Tankha, director of Gangotri Conservation Group, a voluntary organization. Last year

20,000 pilgrims continued on to the glacier. As a result the place is littered with plastic bags, tins and other debris. Husnain counted over 70 'dabhas' or roadside eating places — one close to the glacier's snout — which burn wood and kerosene, releasing smoke and carbon dioxide. Wood burning has led to massive felling of trees.

One likely consequence is change in the microclimate in the region around the



Terminal decline? Gangotri glacier's snout is collapsing.

glacier, say scientists at the university. They are hoping for funds to set up a network of electronic stations to relay weather and pollution data from the area. "The environmental damage can be seen with naked eyes but needs to be studied in detail with instruments," says Husnain. A proposal to establish a Himalayan development authority is already before the Indian cabinet.

K. S. Jayaraman

Staff face pay cuts as CERN prepares to tighten its belt

Munich. Staff at CERN, the European Laboratory for Particle Physics in Geneva, Switzerland, are facing the prospect of a one-year pay cut of 2.5 per cent in exchange for an extra five-and-a-half days leave, as part of a bid by management to reduce running costs.

The proposal is an emergency measure to cope with a major cut in CERN's budget, prompted by Germany's decision to reduce its subscription sharply next year. That decision is expected to be approved at CERN's council meeting tomorrow (20 December).

Staff at CERN have not ruled out industrial action if the proposal, which includes a full review of their pay and conditions next year, is approved. Staff at the European Southern Observatory (ESO) at Garching near Munich are being similarly squeezed, but no cut in basic pay has been proposed.

CERN's council will face various proposals on future costs. The budget scheme most likely to be adopted is that demanded by Germany — a budget reduction of 8.5 per cent for 1997 and 1998, followed by 9.3 per cent for the following two years.

This scheme has the support of some other CERN member states, in particular the United Kingdom. Others, including Italy, are unhappy with the severity of the cuts, but are still likely to vote in favour rather than see Germany pull out of CERN, as it has threatened to do.

Nor do such countries want to defer any budget decision, as that could damage negotiations with non-member countries, such as the United States and Japan, about their hoped-for participation in the Large Hadron Collider programme.

Germany is calling the tune not only on the size of the budget cuts but also on how they should be handled. It believes that pay at the two European laboratories is too generous — a typical researcher may earn around US\$70,000 a year tax-free — and wants much of the proposed saving to be made by reducing personnel costs.

Hans Eschelbacher, German delegate to the CERN council, points out that staff at the laboratory enjoy generous 'expatriate' privileges. All staff hired from outside a 150-km ring around Geneva, for example, receive a supplement equivalent to about 10 per cent of their basic pay. They are also entitled to reimbursement of 75 per cent of their children's private school fees, up to a ceiling of SFr12,000 (\$9,160) a year, and paid home leave every two years.

CERN officials calculate that, even if such subsidies were eliminated overnight — which would be virtually impossible because of contractual obligations — it would save only 6 per cent of total personnel costs, or

about 3 per cent of the laboratory's budget.

But, aware that the writing is on the wall, CERN has negotiated with staff a small reduction in the expatriate supplement. This will fall by half a per cent a year to around 5 per cent for all new staff, beginning in 1997.

CERN's staff association has already accepted this cut. But its officials are angry that management has rejected a proposal to introduce voluntary pay cuts in exchange for extra holiday allowance.

The association will put its case to tomorrow's council meeting. Its deputy president, Derek Ball, says that the association has not decided how it will react if the pay cuts are pushed through. But it will consider taking the issue to the International Labour Organization — also based in Geneva — and may take other forms of industrial action "not excluding strike action".

Some member states share the association's concern about the proposed savings package — but for different reasons. "It is fine for one year, but it is not a long-term solution," says Filippo Menzinger, one of Italy's delegates to CERN council.

He calculates that the unit cost of work will rise because of the reduction in total hours worked per year. The only stable solution, he says, is to work on a rational restructuring of the remuneration system next year.

Meanwhile, ESO, whose council last week approved a German proposal for a five per cent cut in its 1997 budget, has also told its staff that at least two-fifths of the cut will be taken from personnel costs, a compromise on Germany's demand that the entire cut should be absorbed through reduction of staffing costs. The council concluded that the distribution of cuts should be decided by Riccardo Giacconi, ESO's director general.

The council has now set up a working group to propose a new strategy for salary and employment conditions, to be headed by its Dutch delegate, Jan Bezemer.

A spokesman for ESO's staff association, which — unlike its equivalent at CERN — does not have any formal mechanisms for negotiating with management, says that the pressure on personnel costs is unreasonable because staff have already compromised considerably in the last year.

For example, ESO staff have already accepted a reduction of their expatriate supplement from 18 per cent of basic salary to 12 per cent over 12 years for those working in Garching, and from 24 to 18 per cent for those working in Chile. They have also accepted a reduction in the cost of living supplement for those in Chile. And, the spokesman says, some staff positions have already been lost because of internal restructuring during the year. **Alison Abbott**

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

High hopes: Harwell's Oxfordshire site.

Science park future for UK nuclear lab

London. Fifty years after it became the focal point of civil nuclear research in Britain, a converted air force base at Harwell in Oxfordshire is taking on a new role — as the latest UK science and technology park.

The park, to be known as the International Business Centre for Science and Technology, aims to build on Harwell's international reputation to attract high-technology companies to its site 14 miles south of Oxford. "The scientists who arrived in 1946 blazed a trail which has left the area with one of the highest densities of technological brainpower in the world," says Stephen White, Harwell's site manager.

Harwell's transformation into a science park is the latest step in a process that began in the mid-1960s when the research centre, by then part of the United Kingdom Atomic Energy Authority (UKAEA), started to broaden its interests.

The 1990s has seen a gradual transfer of resources to the private sector. Earlier this year saw the privatization of AEA Technology — the body that absorbed much of the commercial and industrial consultancy work carried out at Harwell by the UKAEA.

The site already has several longstanding tenants. These include three Medical Research Council units — which grew out of Harwell's radiobiological work — and the nuclear waste company Nirex.

The presence of such bodies, combined with the UKAEA's remaining research activities, provides "a critical mass which could be the envy of any ordinary business park", says one official. It also means that the plans for the park received enthusiastic support from local communities, concerned about stories of other government centres where privatization has led to significant job losses.

UKAEA officials say that they are negotiating with several potential residents, who will build their own facilities on land that has already been refurbished at a cost of £23 million. The first contracts are likely to be announced in the new year. □

Britain launches two studies of 'Gulf War syndrome' ...

London. Britain's Ministry of Defence (MoD) last week announced two epidemiological studies of illness among Gulf War veterans — two years after the US government set up a study on the possible link between US troops' exposure to chemical weapons during the war and their subsequent illnesses.

The British projects have been selected by the Medical Research Council (MRC), and will be funded by the MoD at a total cost of £1.32 million (US\$2.2 million). Both studies will begin in January, and are expected to take three years.

The first British veterans with symptoms of 'Gulf War syndrome' came forward just months after the conflict ended in 1991. Symptoms range from skin rashes, headaches, nausea and musculoskeletal problems to depression. More than 900 such individuals have been examined under the MoD's 'medical assessment programme'.

So far, however, no evidence has emerged of a new pattern of illness specific to Gulf veterans. The MoD asked the MRC to recommend a direction for future research, for which proposals were invited last May. Two of these proposals have now been put forward by the MRC.

The new epidemiological studies will each compare 3,000 of the 50,000 UK Gulf veterans with 3,000 matched service personnel who did not go to the Gulf. In one study, Nicola Cherry and Alan Silman at the University of Manchester will look for statistical evidence of whether Gulf personnel report more ill-health, and whether they exhibit peculiar combinations of symptoms. "Our initial approach will be to question the troops directly," says Silman.

The second study will compare the reproductive health of Gulf and non-Gulf personnel, along with the health of their children. This work will be led by Patricia Doyle at the London School of Hygiene and Tropical Medicine.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

From soldiers to patients? Fresh studies will assess illness among Gulf War veterans.

If the results do indicate an excess of medical problems in Gulf personnel, then the scientists will need more detailed information about the occupational and environmental exposure of troops to hazardous substances. These include organophosphates (see box, below) and other pesticides sprayed into tents.

Other potential suspects include de-lousing powder used on Iraqi prisoners, vaccinations against transmissible diseases or biological weapons, chemical and biological weapons, and smoke from burning oil refineries, as well as diesel engines in or near tents.

"We are in negotiations with the MoD about what [information] is going to be needed and what access will be allowed," says Silman. The defence ministry has admitted that "medical record-keeping in the Gulf was not satisfactory".

A third study of known health problems and unexplained illness in 3,000 UK veterans and 6,000 controls is being funded by the US government.

Claire O'Brien

...as minister apologizes for 'misleading'

London. Nicholas Soames, the British armed forces minister, told the House of Commons last week that an investigation had revealed how parliament had been repeatedly misled on the question of whether troops had been exposed to organophosphates during the Gulf War.

He apologized for the fact that "flawed advice" originating from the Ministry of Defence had been presented by ministers since July 1994. The admission came during a statement in which Soames announced new research into

the 'Gulf War syndrome'.

Soames first admitted in October that organophosphate pesticides had been widely used to prevent troops from contracting flyborne diseases — in contrast to previous government statements that use of organophosphates was minimal.

The new information could be critical to the success of the new research, said Soames, because many of the health problems reported by Gulf War veterans could have been caused by exposure to organophosphates.

C. O'Brien

Advances encourage Nobel winner to head AIDS vaccine effort

Washington. David Baltimore, the Nobel prizewinning biologist, said last week that the promise of recent advances in AIDS vaccine research had helped to persuade him to agree to lead the US National Institutes of Health (NIH)'s \$129-million vaccine research effort.

Baltimore, professor of molecular biology and immunology at the Massachusetts Institute of Technology, says that he decided to accept the job after launching his own investigation of developments in the field. "I wasn't going into something where it was my belief that it wasn't going to work. And I must say that I started off with some doubt," he says. What he found, he adds, "[has] given me a lot of hope that it is possible" to make a vaccine.

Among the advances that Baltimore said spurred his optimism were the discovery of the chemokine receptor, and the demonstration that monkeys can be protected with a live attenuated virus.

Announcing Baltimore's appointment last week, NIH officials said his gifts and experience are perfectly suited to the formidable task. "I'm very pleased," said Harold Varmus, director of NIH. "[David] has got a lot of exciting ideas. This dovetails ...with our determination to do something about HIV vaccines in the near future."

William Paul, director of the Office of AIDS Research, said that Baltimore's creativity, as well as his background as both an eminent virologist and molecular immunologist, will provide the kind of "visionary" leadership needed to develop a vaccine.

Working as a consultant to the National Institute of Allergy and Infectious Diseases (NIAID), Baltimore will oversee vaccine research across the whole of NIH. He will also head an AIDS Vaccine Research Committee, which will be composed mainly of non-government scientists.

Baltimore, who was a co-discoverer of reverse transcriptase, and has done leading work on the polio virus, will spend about 20 per cent of his time coordinating the NIH's vaccine research. His appointment is a response to a recommendation of a review of AIDS research at NIH published last March by a working group led by Arnold Levine, a molecular biologist at Princeton University.

Meredith Wadman

See also: *AIDS research strategy* (page 606)

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Baltimore: "visionary" leadership for task.

Margaret Lampert

Improbable museum seeks irrelevant items

Boston. A towel that may have been used by Albert Einstein and freeze-dried ice cream similar to that consumed by US astronauts are among the bizarre exhibits in Boston's new Museum of Improbable Research.

The museum, at Harvard University, opened earlier this month with little fanfare. "This is the best-kept secret on the Harvard campus," says Marc Abrahams, the museum's curator and editor of the affiliated journal, *Annals of Improbable Research* (AIR).

Still in its formative stages, the museum displays an odd assortment of items. Prize exhibits include remnants of moulds used to make plaster casts of the feet of Nobel prizewinning scientists, and notes taken from laboratories in which famous scientists worked.

Other exhibits include an unclaimed 1996 Ig-Nobel prize and a male pin-up 'Studmuffins of Science' calendar. There is also a reproduction of the original working papers of a scientist whose taxonomical analysis proved that a popular television figure, Barney from the *Flintstones*, is not a dinosaur after all, and a glass jar containing a Barney specimen bathed in formaldehyde.

Abrahams hopes that the collection will expand, and is encouraging people to send 'irrelevant objects' either to him or to his associates, the Harvard chemists Dudley Herschbach and William Lipscomb.

"We're looking for something different — not the kind of stuff you can see in a normal science museum, buy in a normal science museum, or steal from a normal science museum," Abrahams explains. "We specialize in the work in progress that was cut off before there was any progress." Cold fusion apparatus is one of many items on his wish-list.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Research apparatus from a failed experiment on anorexia nervosa from the museum.

Alice Shirrell Kaswell

"The idea is to draw attention to, or to commemorate, research efforts that are unlikely to receive funding through normal channels. We hope that, as the collection grows, some objects that eventually served as the impetus for a great scientific discovery will be donated because people laughed at them at the time."

The Museum of Improbable Research is housed on the top floor of Harvard University's Conant Laboratory in an abandoned greenhouse that looks out over the Boston skyline. The greenhouse is unique, says James Michel, a molecular biologist at the university who is also an associate curator. "Too hot in the summer, too cold in the winter, this spot has its own special thermodynamic qualities."

Others have welcomed the latest arrival on the New England museum scene. Jerry Reilly, executive director of the Boston-based Museum of Bad Art, says of his Harvard counterpart: "We wholeheartedly support the important work that they're doing. The arts and sciences have been separated for too long. We're hoping that these two facilities can help bridge the gap between the bad art and bad science worlds." Steve Nadis

Hopes rise for agreement on greenhouse gas targets

Washington. A consensus appears to be growing steadily among signatories to the international climate convention of 1992 on the need for binding targets for emissions of greenhouse gases. Although a review meeting in Geneva last week made little progress towards reaching an agreement on targets, there was little dissent from the principle that they are needed.

That encouraged diplomats, who hope to agree a treaty on such targets at a summit in Kyoto, Japan, next December. But environmental groups still accuse developed countries of foot-dragging on the issue.

The fifth meeting of the Ad-hoc Group for the Berlin Mandate, the body that is supposed to draft the treaty, was used by the leading industrial nations to put forward their various proposals for emissions limits. The United States set the tone with a conservative proposal document which said that 2010 should be the earliest date for binding limits on emissions, and suggested considerable flexibility in the way targets are set.

The United States proposes, for example, that countries that are unable to meet their immediate targets for one period should be allowed to 'borrow' part of their allowance from the succeeding period.

At a meeting of environmental ministers at the beginning of last week, the member states of the European Union (EU) failed to back a proposal to cut emissions by 5 to 10

per cent by the year 2005, and 15 to 20 per cent by 2010. As a result, at the Geneva meeting the EU could only repeat its previous support for binding targets at some date after 2000, without adding any specifics.

Japan and France — the latter breaking ranks with its EU partners — each proposed that any targets eventually agreed should be based on a maximum volume of emissions per capita. Critics described the proposal as self-serving, as both countries have lower emissions per person than other industrialized countries as a result of their heavy reliance on nuclear power.

Environmental groups had hoped that the meeting would move closer to agreement on the type of mandatory limits that could be included in a new treaty. Dan Lashof of the Natural Resources Defense Council, a Washington-based environmental group, says: "The United States is backtracking, and none of the other major players are ready to fill in for them."

But Eileen Claussen, assistant secretary at the US State Department, argues that the meeting had several positive aspects. "It's the first meeting where there hasn't been any contentious discussion about the science" of global warming, she says, adding that "everyone seemed to accept the idea of binding targets" for industrialized countries.

Claussen agrees that the US proposal to allow "borrowing" of emissions "made a lot

of people nervous", and says that "we'll have to reassess it". She says any agreement reached at Kyoto will require enabling legislation, as well as ratification by the US Senate.

The lack of a strong lead at Geneva from either Europe or Japan makes the US position crucial in the preparations for Kyoto, say observers. But US officials say that they will not decide on Washington's final position until a late stage.

Since announcing its support for mandatory targets in July (see *Nature* 382, 287; 1996), the Clinton administration has come under pressure from industrialists and Republican leaders in Congress, and is now seeking targets that would be sufficiently relaxed to allow a gradual transition. "If we negotiate something that upsets corporate America, the chances of ratification are zero," says one administration official.

The *ad hoc* group will meet three more times before the full meeting of the Conference of the Parties at Kyoto next December. Despite pressure from environmental groups, early drafts of a treaty are not expected to appear until next June.

At further meetings in Geneva this week, the subsidiary body on Scientific and Technological Advice was due to address the role of aircraft emissions and to decide what input the Intergovernmental Panel on Climate Change can make to preparations for the Kyoto meeting. Colin Macilwain

Priorities for AIDS research set out by US task force

Washington. A US interagency task force this week released a 40-page strategy to guide the AIDS policy of the Clinton administration for the coming four years. The main goals include research to develop more effective treatments, a preventive vaccine, and a cure for AIDS. Another goal is to ensure that research is translated into improved prevention and enhanced care.

The report, which has taken a year to prepare, calls for the development of microbicides, urges further work on protease inhibitors, and endorses a vaccine strategy for the National Institutes of Health (NIH). President Bill Clinton was expected to accept the report on Tuesday (17 December).

"It's the first time that the US government has developed a comprehensive strategy covering all federal agencies and providing a blueprint for the years ahead," said Richard Sorian, a spokesman for the White House Office of National AIDS Policy. The report endorses the continued oversight of all NIH AIDS research by the Office of AIDS Research. □

New hope on landmines

Boston. Possible strategies for reducing the threat to civilians posed by the estimated 120 million landmines buried in 62 countries — as well as the millions of unexploded 'bomblets' scattered throughout Laos — were published last week. They are based on a week-long 'brainstorming' workshop held at Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts, last August.

A panel of scientists from MIT, Harvard University, Los Alamos National Laboratory, the Naval Research Laboratory, IBM and elsewhere recommended an array of new technologies. These include a

sensor called the Meandering Winding Magnetometer and various robotic devices, and could accelerate the pace and enhance the safety of demining activities by a factor of 100 to 1,000 within five years. "We went into this workshop feeling that little could be done, but came out feeling that a great deal can be done," says Kosta Tsipis, director of MIT's Program in Science & Technology for International Security and the author of the report. □

Access to NIH clinical centres

Washington. A US National Institutes of Health (NIH) panel assessing the agency's clinical research is expected to recommend that 76 NIH-supported clinical research centres should open their doors to investigators not supported by the NIH. The recommendation is contained in a report due for release shortly by the NIH Director's Panel on Clinical Research, chaired by David Nathan, president of the Dana-Farber Cancer Institute in Boston.

Nathan says the 15-member panel is seeking "the broadest possible access of good ideas" for the centres, known as General Clinical Research Centers. "We don't want anybody excluded because they don't happen to get their grants from the NIH." The centres, which each receive an average of US\$2 million in NIH funding, support scores of NIH-backed investigators nationwide. Under the new proposal, the centres would be opened to peer-reviewed clinical investigators supported by industry and private foundations. □

Russian researchers protest

Moscow. More than a hundred nuclear researchers — previously the most favoured among Russian scientists — picketed the Council of Ministers and the finance ministry in Moscow last week to demand more funds for research. Last year, the 30 nuclear institutes that account for the core of Russia's military research received only two-thirds of their promised US\$900-million budget, and plans

New **Swedish** solutions for purifying peptides
of **any source** using **any technique**



INTRODUCING:

- ten new reversed phase chromatography columns
- two new ion exchange chromatography columns
- a new size exclusion chromatography column
- one completely new system for peptide, oligonucleotides and other biomolecules (ÄKTA is the Swedish word for real, it's pronounced eckta).

being discussed in the State Duma (the lower chamber of the Russian parliament) would reduce this to \$362 million next year. Trade union leaders who organized the protest say that employees at some institutes have not been paid for several months. □

Hobby-Eberly telescope sees light

Washington. A new 11-metre telescope at the McDonald Observatory in western Texas passed an important milestone last week by seeing its 'first light'. Following an engineering check-out period, the \$13.5 million Hobby-Eberly Telescope (HET) will begin full science operations late next year. The telescope's relatively low cost — it is only about 15 per cent of the cost of comparably-sized instruments, such as the Keck Telescope in Hawaii — is due to its being limited to spectroscopy and only viewing 70 per cent of the sky. Principal partners in the project are the University of Texas and Pennsylvania State University. □

ITER plasma model queried

Munich. Questions have been raised about the validity of a theoretical model predicting that higher-than-anticipated turbulence in the proposed International Thermonuclear Experimental Reactor (ITER)'s fusion plasma will effectively prevent it from reaching and maintaining a temperature high enough to ignite a fusion burn. The challenge came last week in a statement from an expert group that includes representatives from each of the four international partners participating in plans for ITER.

The model, revealed two weeks ago in the journal *Science*, has been developed by William Dorland and Michael Krotschenreuther from the University of Texas in Austin. Proponents are calling for its predictions to be taken on board by ITER's design teams. But Karl Lackner of the Institute for Plasma Physics in Garching, Germany, says that the model has already been discussed widely within the

fusion community. And the expert group has issued a statement describing the Dorland and Krotschenreuther model as having only an "average performance" in reproducing results of current experiments compared to other models that give more favourable predictions for ITER. □

Genentech co-founder quits

San Francisco. Robert Swanson, the California investment banker who helped to launch the biotechnology industry in the mid-1970s, announced last week that he is to resign as chairman of Genentech at the end of the month. Swanson co-founded the San Francisco company in 1976 with Herbert Boyer, professor of biochemistry at the University of California, San Francisco. Swanson, who is 49, says that he is leaving to concentrate on his "first love" — financing innovative young companies. □

Indian physicist awarded prize

New Delhi. **The Third World Academy of Sciences has named the Indian physicist Mambillikalathil Govind Kumar Menon (right) as the winner of the second Abdus Salam prize for science and technology, the academy's highest award. The first award was made to Federico Mayor, the director-general of Unesco.**

Menon has held several leading positions in the Indian government in policy-making and administration, and was at one time minister for science. The academy, based in Trieste, Italy, says that the award is being made for his "distinguished contributions to cosmic ray and high-energy particle physics and to the development of science in the Third World". □



Are you working with natural peptides? Synthetic peptides? Recombinant peptides? Peptide fragments? Or all of them? Whatever peptide you work with, your options for purifying them just increased.

With ten new columns for reversed phase chromatography, you're nearly certain of finding the selectivity you need in our extensive range. All of these new RPC columns deliver high resolution; combined, they'll take you from purification and analysis to peptide mapping.

We can support you with advice and solutions for other peptide purification techniques as well. Are you separating peptides with poor solubility? Our new size exclusion column withstands high pH and solvents. Do you need an extra technique to help you with a difficult-to-separate peptide? We have two new ion exchange columns that permit very high resolution and withstand high pH ranges.

What's more, all 13 columns are supported by ÄKTA™ purifier—a revolutionary new purification system for peptides, oligonucleotides and other biomolecules. Its preset protocols let you resolve all major purification tasks quickly and easily. Its control system lets you instantly transfer your methods to purification systems at all scales.

Want to know more about our peptide purification solutions? Call us: 1 (800) 526 3593 from the USA; +81 (0)3 3492 6949 from Japan; or +46 (0)18 16 50 11 from Europe and the rest of the world; or meet us on the Internet at <http://www.biotech.pharmacia.se>.



**Pharmacia
Biotech**
Uppsala, Sweden. (And the rest of the world)

Beyond the language barrier

SIR — A recent note in *Nature*¹ tells of the Japanese chemist Eiji Osawa who was devastated when he first saw the original 1985 paper describing the discovery of buckminsterfullerene. In 1970, while at Kyoto University, Osawa had predicted the possible existence of a C₆₀ molecule with a geodesic structure, which he published in Japanese. Fifteen years later his work was unknown to the group that subsequently received the Nobel prize².

Ignorance of published work within various fields is still a major problem. Abstracting and indexing services continue to improve and provide information on the latest works published, but the proliferation of international and regional journals makes it difficult to keep up with the literature. It is understandable that many English-speaking scientists fail to read journals in foreign languages, such as Japanese, without translations. But ignorance of important articles in local foreign journals published in English is also increasing.

The Japanese have made an active effort to publish society and local journals in English. Within the field of neuroscience, journals such as *Neuroscience Research* (from the Japan Neuroscience Society) and *Journal of Brain Science* (Japan Brain Science Society) are examples of important English-language journals based in Japan. And many of the 80 Japanese medical schools publish their own English-language journals. With an increasing number of countries with important research enterprises, it is clearly difficult to keep up with discoveries by international scientists unless they are

published in international journals. But there are many reasons why a scientist may want to publish important work in a particular local or regional journal.

Access to scientific information is important. While it could be argued that scientists should try to publish their most important work in leading international journals, that is unfair. The burden is on each of us to find innovative ways to keep in touch with the ever-expanding scientific literature. Abstracting, indexing and Internet services should continue extending their coverage of the world's scientific journals. If it is assumed that the international scientific language is English, then regional and local foreign institutions should continue Japan's approach of providing important English-language journals. In this way important discoveries such as the C₆₀ carbon atom, as originally described by Osawa, will not be rediscovered 15 years later because of a language barrier.

Paul R. Sanberg
Cesario V. Borlongan
Hitoo Nishino

*Division of Neurological Surgery,
University of South Florida
College of Medicine,
12901 Bruce B. Downs Blvd,
Tampa, Florida 33612, USA
and
Department of Physiology,
Nagoya City University Medical School,
Mizuho-ku,
Nagoya 467, Japan*

1. Swinbanks, D. *Nature* **383**, 562 (1996).

2. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318** 167–163 (1985).

Closure explained

SIR — I was disappointed to read your recent News article "Closure of astrophysics laboratory blamed on Enola Gay 'fallout'" (*Nature* **383**, 367; 1996). The author's suggestion that closure of the Laboratory for Astrophysics is some form of "retribution" against a former museum director are supported with a single quotation from one anonymous "colleague" of a staff member. Allow me to set the record straight.

The Laboratory for Astrophysics is being closed because current fiscal realities dictate that the National Air and Space Museum must concentrate on its core function, the care of the national collection. A report in October 1995 by the General Accounting Office describes in great detail the critical needs of the air and space collection. Similarly, the September 1995 management review by the National Academy of Public Administration concludes that the "research activities pursued by the astrophysics laboratory are not a crucial element of the museum".

Although the primary funding for the laboratory came from grants, two of the three employees' salaries came from federal funds. Closing the laboratory is one of many difficult decisions that the museum's management has had to make to ensure that the museum continues to meet its primary responsibility, stewardship of an unparalleled collection of unique artefacts from aviation and space exploration history. To suggest any other motivation does a disservice to the entire staff of the museum.

Donald S. Lopez
*National Air and Space Museum,
Smithsonian Institution,
Washington, DC 20560, USA*

Genome patents

SIR — Any effort to illuminate the debate on the ownership of the human genome by an objective quantified analysis is laudable. But failing to take into account the pace at which techniques and methods evolve can greatly limit the scope of the results. The recent analysis by members of the Science Policy Research Unit (SPRU) at the University of Sussex in England (*Nature* **380**, 387–388; 1996) focused on human DNA sequence patents granted before 1995, and so covered work carried out mainly with classical techniques. We have adopted an alternative approach, surveying filed (and not only granted) patents for human and non-human sequences, as well as methods of genome analysis.

Our results draw on two databases: the *Science Citation Index* for scientific research at the international level (50,903 bibliographic references for 1991–95) and the

Derwent Biotechnology Abstracts for patents (6,105 patents for the same period — five times more than the SPRU database).

Like the SPRU team, we found that Europe was lacking in capacity to transfer basic research to industrial applications. Europe accounts for 39% of all scientific publications on genomic research but is responsible for only 21% of all patents. This weakness, which is not limited to genomic research, will require specific measures if Europe is to remain competitive.

The main difference between our results and those of the SPRU team lies in the private sector. Analysing the most involved organizations, we found that the proportion of Japanese companies filing patents (37%) is roughly the same as that of US companies (35%), whereas the SPRU team found that Japanese companies owned 50% of patents and that US companies owned 25%. And although public-sector research is marginal in the SPRU analysis, with public institutions owning 17% of patents granted world-

wide, it is central in our results, accounting for 38% of all patents filed (United States 54%, Europe 43%, Japan 0%).

Overall, this puts the United States in the lead (47%) with Japan (26.6%) a serious competitor to Europe (21%). Which results should be recognized? It depends on whether one wants to identify the current owners of genes or the players developing the competence and tools that give them a good chance of owning such patents tomorrow. Bearing in mind that only 1 per cent of all genes are today protected by patents, an analysis based on the identification of potential and not only of acquisition has clear advantages.

P. B. Joly
M. A. de Looze
V. Mangematin
*INRA/SERD,
R. Coronini*
*IREPD,
Université Pierre Mendès-France,
BP47, 38040 Grenoble Cedex 09, France*

Smokescreen lifted on insect flight

R. McNeill Alexander

The conventional laws of aerodynamics cannot explain how insects fly. But work using a mechanical model shows which of several possible unconventional mechanisms generates enough lift for an insect to stay airborne.

INSECT flight cannot be explained by the conventional aerodynamic principles that are used by aircraft engineers. Many suggestions have been made for unconventional mechanisms that might explain insect motion. But, on page 626 of this issue, Ellington *et al.*¹ give us the first clear evidence that one particular mechanism is the important one. In effect, they have discovered how a typical insect flies.

Conventional aerodynamics is concerned mainly with steady motions, such as a wing moving at constant velocity or a propeller that rotates at a constant rate. In contrast, the flapping motion of an insect's wing is highly unsteady — it keeps changing direction and speed. Upstrokes alternate with downstrokes, tens or even hundreds of times per second, and at each change of stroke the wing rotates about its long axis, tilting to the appropriate angle for its new direction of motion.

In previous attempts to explain insect flight by conventional aerodynamics, bold liberties were taken with the physics. A flying insect or an aircraft is supported by lift forces on its wings. We knew how much lift could be produced by wings of given size, moving steadily at a given speed. But we assumed that, at each instant during the insect wing-beat cycle, the lift forces were the same as if the wings were moving steadily with the same velocities. Moreover, we imagined each wing could be divided into thin slices — rapidly moving slices near the tip of the wing and more slowly moving slices near the base — and for each instant the force on each slice was calculated as if it were part of a long wing that was moving at a uniform velocity. In retrospect, it was remarkably optimistic to have expected that this approach would work, but it seemed to be worth trying.

In some cases (for example, locusts²), the 'quasi-steady' approach appeared to explain flight adequately. But in others

(for example, the tiny wasp *Encarsia*³), the predicted lift forces were much too small to keep the insect airborne, and various 'unsteady' effects were suggested. The most famous and most convincing of these was the 'clap and fling mechanism', proposed by Torkel Weis-Fogh³. This theory seemed to explain how *Encarsia*, among others, flies. These insects clap their wings together at the top of the upstroke, and Weis-Fogh showed how the subsequent

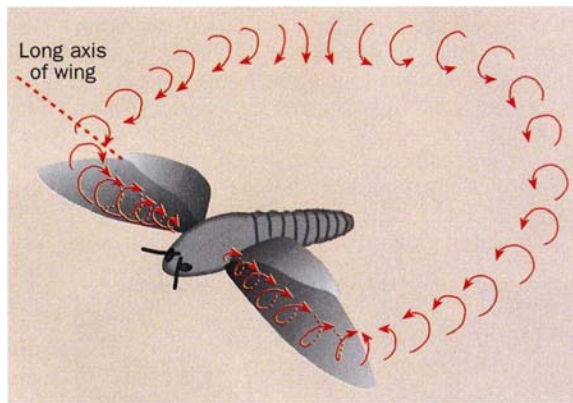
These adjustments satisfy the criteria for generating a similar pattern of airflow.

To gain lift, air must be driven downwards and, as this happens, vortices will be formed around the downward-moving jet of air. Vortices are cylinders of air that are caused to rotate by a puff of moving air, like rollers that rotate when a crate is pushed over them. Photographs of the airflow showed that, with each downstroke of the wings, vortices form over the leading edges (see figure), and they extend back from the wing tips to form a ring around the puff of air that is generated by the stroke. These vortices are the visible sign of the all-important downward air movement.

The experiment with the real moth gave a strong indication that leading-edge vortices were present, but only the mechanical model could show the airflow within them. The model showed that air in these vortices does not simply revolve in circles, but that it flows out towards the wing tip in ever-widening helices. This could not have been predicted by conventional aerodynamic analysis, as by dividing the wing into slices airflow was effectively restricted to two dimensions.

It was already known that leading-edge vortices can be generated in two ways: by rotation of the wing about its long axis as it is tilted to the appropriate angle at the start of each stroke; and by translation (that is, forward movement) of the wing during the stroke. A leading-edge vortex will persist during steady translation, but the phenomenon known as 'delayed stall' enables the vortex to be stronger and the lift to be greater shortly after motion starts. If a wing is tilted too much in an attempt to get more lift, it will stall and therefore lose lift instead of increasing it. But the stalling is not immediate and, until it happens, increased lift is possible.

A vortex resulting from rotation of the wing will already be at full strength when translation starts at the beginning of the downstroke. This vortex will be strongest where the wings are widest — at the base. But a vortex that is due to translation will grow in the early stages of the downstroke, and be strongest at a point on the wings midway between the base and the tip. The experiments with the model



Vortices generated by the wing movements of a flying hawkmoth. Ellington *et al.*¹ have used a mechanical model to show that the leading-edge vortices generated by flapping of the insect's wings are stabilized by a phenomenon known as 'delayed stall', which allows the insect to generate enough lift to stay airborne.

separation of the wings can generate increased lift forces. Meanwhile, more and more cases were found in which conventional steady aerodynamics failed (reviewed in ref. 4).

Our understanding of bird flight was enhanced by research that allowed us to see movements of the air in the wake of flying birds⁵. But to discover the mechanisms that generate the forces required to stay airborne, we needed to look, not at the wake, but at the airflow around the wings. This is what Ellington and colleagues¹ have done. In one experiment they photographed streaks of smoke passing over the wings of tethered hawkmoths that were flapping in a stream of moving air. The experiment simulated the flight of a free moth in still air; however, it failed to show the fine details of the airflow. So the authors constructed an elaborate mechanical model that simulated the moth's wing motion. This had ten times the length and breadth of the moth, and it operated at one-hundredth of the frequency.

1. Ellington, C. P., van den Berg, C., Willmott, A. P. & Thomas, A. L. R. *Nature* **384**, 626–630 (1996).
2. Jensen, M. *Phil. Trans. R. Soc. Lond. B* **239**, 511–552 (1956).
3. Weis-Fogh, T. *J. Exp. Biol.* **59**, 169–230 (1973).
4. Ellington, C. P. *Symp. Soc. Exp. Biol.* **49**, 109–129 (1995).
5. Rayner, J. M. V. *Symp. Soc. Exp. Biol.* **49**, 131–155 (1995).

showed that, by both criteria, the vortex that persists through the downstroke is due to translation. So the unsteady effect that makes flight possible must be the one that depends on translation; that is, delayed stall.

To give more lift than would be possible in steady motion, the vortices must be stronger than conventional aerodynamics would predict could persist. Ellington and colleagues suggest that the unexpected helical flow — out towards the wing tips —

may help to stabilize the vortex, enabling it to retain its early strength for longer than would otherwise be possible. It remains to be discovered whether similar effects occur in the flapping flight of birds. In some cases, such as in slow or hovering flight, birds also generate more lift than conventional aerodynamics can explain. □

R. McNeill Alexander is in the Department of Biology, University of Leeds, Leeds LS2 9JT, UK.

OPTICAL STORAGE

The future is looking blue

Pauline Rigby

COMPACT disks in their existing form may soon be superseded. Current CD technology uses near-infrared laser diodes to read information, because they are long-lived and reliable. But blue would be much better than red. Because the size of a focused spot scales as the square of the wavelength, reducing the wavelength by a factor of two would increase the storage capacity by a factor of four. No wonder there is a concerted effort to produce a current-injected blue laser diode for continuous wave (c.w.) operation at room temperature. Last

inversely related to band-gap energy, shorter wavelengths must come from materials with larger band gaps. The two candidates for providing shorter wavelengths are classes of material based on either gallium nitride (GaN) or zinc selenide (ZnSe). Work on ZnSe lasers has been in progress for longer, and working laser designs using this material are more sophisticated², but GaN may offer better prospects in the long term.

For a start, the band gaps available from GaN-based materials range from

tion of a 514.7-nm laser⁵. One hundred hours in the laboratory is a commendable duration, but still nowhere near good enough to put in a piece of equipment and sell. For that, lifetimes of the order of 10,000 hours are needed.

Increased lifetimes have been brought about by improving material quality — but even a perfect crystal of ZnSe will be susceptible to runaway defect generation, the process that kills lasing action. The quality of ZnSe being produced now is extremely high, so can much more performance be squeezed out of this material? In contrast, GaN is exceptionally stable at high temperature and under chemical attack, and so it should be resistant to the generation of defects.

Despite their immunity from the problem of defect multiplication, GaN lasers also fail in a very short time, because they get too hot. As the device heats up during c.w. operation, the difference in expansion between layers of different materials in the device causes it to break up under the strain. Some strain is already present before the device is switched on, because the layers have to be grown on a substrate material whose lattice does not match GaN, and this exacerbates the problem. Single-crystal bulk GaN would be the ideal choice of substrate material — then the crystal lattices would match exactly — but none is available at present.

Nichia's laser diode survived for less than one second when operated continuously at room temperature. By cooling it only a little to 243 K (−40 °C), however, its lifetime was increased to more than 30 minutes, long enough to allow spectral measurements. Spectra had been obtained before with pulsed current, which allows heat to dissipate between pulses. These spectra contained several sharp peaks, due to the temperature changes in the device during pulsed operation, but the c.w. results show a single sharp emission peak — proof of lasing action.

The design has already been improved — at the 1996 meeting of the IEEE Lasers and Electro-Optics Society, Nichia announced continuous operation at room temperature for 27 hours. It seems likely that they will meet their target for commercial production of blue laser diodes within two years. That's lucky, because systems are already being designed that will use the new blue laser technology. □

Pauline Rigby is in the Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.



Blue-chip research. Gallium-nitride-based blue laser diodes such as this may soon be in our CD players and multimedia machines. (Picture courtesy of Nichia Chemical Industries Ltd.)

month in *Applied Physics Letters*, Nakamura *et al.* of Nichia Chemical Industries reported the first successful c.w. operation of a blue laser diode made from gallium-nitride-based materials¹.

A group of Japanese companies, headed by Toshiba and Matsushita, is working towards a common specification for the new 'digital video disk', which will have a data capacity of 4.7 gigabytes per side, 12 times as much as an ordinary CD. Only a 50 per cent increase in capacity is due to the reduction in wavelength (from the 780 nm used now to a red laser operating at 635 nm); the rest of the improvement is obtained by increasing the numerical aperture, which reduces the spot size of the focused laser beam, and by improving data compression algorithms. But with a blue laser, a digital video disk system could have a total capacity of 15 gigabytes.

Near-infrared lasers are based on gallium arsenide, which has a band-gap energy of 1.4 eV. Because wavelength is

1.9 eV up to 6.2 eV, whereas ZnSe has a gap of 2.67 eV, so GaN materials are potentially capable of providing much shorter wavelengths, extending into the ultraviolet. The shortest-ever semiconductor laser wavelength, 376 nm, was obtained from a GaN diode³.

Nichia's new blue laser⁴ uses four Ga_{0.2}In_{0.8}N quantum wells for light emission at 411 nm. Quantum wells are layers of material only a few tens of atom layers thick, and they are more efficient at producing light than is bulk material because they attract electrons and holes (which recombine to produce photons).

Second, defects multiply easily in ZnSe lasers². Defects act as nonradiative sites — places where the electric current is converted into lattice vibrations instead of photons — and these show up as dark spots on the device.

The best performance from a ZnSe laser so far was reported by Sony, who demonstrated a lifetime in excess of 100 hours for room-temperature c.w. opera-

1. Nakamura, S. *et al.* *Appl. Phys. Lett.* **69**, 3034–3036 (1996).

2. Morkoc, H. *et al.* *J. Appl. Phys.* **76**, 1363–1398 (1994).

3. Akasaki, I. *et al.* *Elec. Lett.* **32**, 1105–1106 (1996).

4. Nakamura, S. *et al.* *Appl. Phys. Lett.* **69**, 1477–1479 (1996).

5. Taniguchi, S. *et al.* *Elec. Lett.* **32**, 552–553 (1996).

Wet transport proteins

Jared M. Diamond

WATER transport is a hallmark of epithelial sheets of cells, such as those of the intestine, kidney and stomach. As well as being physiologically important, this process is clinically relevant in oral rehydration therapy of patients suffering from cholera and other bacterial diarrhoeas. Although the usual theories of epithelial water transport have invoked distinctive features of epithelial ultrastructure, Loo *et al.*¹ have now demonstrated massive coupling of water flow to solute transport when an intestinal glucose transporter is expressed in a cell devoid of epithelial ultrastructure. This observation implies that the transporter itself is co-transporting water — a remarkable finding, which bears on our understanding of transport proteins as well as of epithelia.

It has long been known that epithelial absorption or secretion of water is somehow coupled to concurrent active transport of solutes. The fact that, in many epithelia, water and solutes move in proportions that are isosmotic to the bathing solution or blood suggested that the coupling mechanism involves osmosis. By about 30 years ago, attention turned to hypotheses invoking local osmotic gradients established within epithelia by active solute transport, and made possible by the role of infolded epithelial ultrastructure in retarding diffusional solute loss^{2,3}. But since then, attempts to detect the postulated small osmotic gradients have been unsuccessful, encouraging the search for alternative hypotheses.

Loo and colleagues' approach was to inject complementary RNA of the rabbit intestinal brush-border Na⁺/glucose co-transporter SGLT1 into *Xenopus* oocytes, where it becomes expressed at high levels of more than 10¹¹ copies in each cell. Coupled co-transport of two Na⁺ ions and one glucose molecule was measured electrophysiologically as the transmembrane electric current associated with Na⁺ movement, while water transport into the oocyte was calculated from changes in cell volume, measured by photographing the oocyte at one-second intervals. The rate of Na⁺/glucose co-transport was manipulated in four ways: by adding sugar substrate, adding the transport inhibitor phlorizin, varying the temperature or varying the membrane potential.

It turned out that activation of Na⁺/glucose co-transport into the oocyte is associated, within 10 seconds, with entry of water. Phlorizin abolishes both the Na⁺-related current and water entry. The activation energies for Na⁺/glucose co-transport and for water entry are virtually the same, about 25 kcal mol⁻¹ — much higher than the values of 3–9 kcal mol⁻¹ observed for pore-forming water channels

in membranes. Thus, SGLT1 co-transport water along with Na⁺ and glucose. Assuming that each transporter cycle conveys one sugar molecule and two Na⁺ ions, plus (for electroneutrality) two anions, then the observed stoichiometry of 260 mols H₂O per transport cycle suggests that 87 mols H₂O per mol of solute are conveyed by SGLT1 itself.

To those trained to associate solute-coupled water transport with the unique ultrastructure of epithelia, this observation

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Close-up on the intestine. Light micrograph showing the highly folded nature of the intestinal epithelium, which has the brush-border membrane as its outer layer.

of such transport in an isolated simple cell is astonishing. However, like the apocryphal reaction to Columbus's boast that he could make an egg stand on its end (as he then showed, by propping up the egg's base with a pile of salt), our reaction to epithelial water transport without epithelia becomes "of course" when we hear the solution. As Loo *et al.* point out, SGLT1 can bind large glucosides of molecular volume up to 250 Å³, so the presence of hundreds of water molecules around SGLT1's sugar-binding site is reasonable. X-ray crystallographic studies of sugar-binding proteins reveal that water occupies up to 10 per cent of their volume, and sugar binding can modify that amount of bound water. So it would not be surprising for a conformational change in SGLT1, associated with sugar binding, to lead to water transport by causing a large change in water binding during each transport cycle. In support of this interpretation, Loo *et al.* mention findings or preliminary observations of water co-transport with six other ion co-transporters. The postulated phenomenon may thus be of very widespread significance — even in plants.

The new work throws up at least four questions, to some of which the authors themselves suggest possible answers. First, epithelia consist of two cell membranes in series, the so-called brush-border membrane and the basolateral membrane. SGLT1 is confined to the brush-border

membrane, raising the issue of how water co-transported across that membrane with SGLT1 then passes through the basolateral membrane and into the bloodstream. But the basolateral membrane also contains transporters and ion co-transporters, including a different glucose transporter (GLUT2); perhaps a related water co-transport process associated with one of those transporters is what carries water across the basolateral membrane.

Second, die-hard supporters (such as myself) of osmotic mechanisms for epithelial water transport will scrutinize the arguments by which Loo *et al.* dismiss osmotic interpretations of their observations. The authors point out that, in principle, SGLT1-driven solute transport into the cell would tend to raise intracellular osmolarity and carry water into the cell osmotically, even without any water co-transport by SGLT1 itself. But, given the low measured osmotic water permeability of the oocyte, Loo *et al.* calculate that the osmotic inflow of water would be less than 1 per cent of the observed inflow if the cell interior is well mixed, and still only 3 per cent of the observed inflow if one takes account

of unstirred-layer effects (calculated by an equation assuming a flat membrane).

However, the oocyte surface, like an epithelium itself, must be highly infolded, because Loo *et al.* report that its effective surface area (calculated from capacitance measurements) implies that there is enough infolding to increase the area by nine times. In the presence of membrane folds, solute transport becomes much more effective at generating water flow osmotically, and this fact was the basis of theories attempting to account for epithelial water flow by local osmotic effects. Could the oocyte actually be operating like a miniaturized epithelium, with its own membrane folds increasing osmotic coupling of water flow to solute transport? Calculations based on the actual geometry of the oocyte membrane will be necessary to resolve this question.

Third, solute movements in a voltage-clamped oocyte are complicated. In addition to the two Na⁺ ions and one glucose molecule conveyed by SGLT1, current is being carried by Na⁺ across the oocyte membrane, and ions are also being delivered directly into the cell by the intracellular voltage-clamping electrode as required by electroneutrality. Those non-SGLT1-related ion movements will tend to be associated with water flow through the transport-number effect⁴. Without a detailed analysis, it is unclear how much of the observed total water

Marshall Sklar/SPL

flow is directly associated with the co-transporter, and whether the number of water molecules moving in association with each SGLT1 cycle is fewer or greater than 260.

Finally, some may still wonder how to account for the observation that solute-coupled water transport by the small intestine, gall bladder and kidney proximal tubular epithelium is isosmotic to the bathing solution within experimental error. On the face of it, this is difficult to reconcile with any hypothesis of epithelial water transport that does not involve osmotic equilibration. The apparent coupling ratio for SGLT1 operation — 87 H₂O molecules per solute molecule — is far from isosmotic to the bathing solution used by Loo *et al.* (about 200 milliosmolar, or 275 H₂O

molecules per solute molecule). Could co-transporters be contributing much water flow while osmotic mechanisms account for the final osmotic equilibration? Questions such as these will keep epitheliologists, as well as membrane biologists and protein chemists who have no particular interest in epithelia, busy for some time. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90095-1751, USA.

1. Loo, D. D. F., Zeuthen, T., Chandy, G. & Wright, E. M. *Proc. Natl Acad. Sci. USA* **93**, 13367–13370 (1996).
2. Curran, P. & MacIntosh, J. *Nature* **193**, 347–348 (1962).
3. Diamond, J. & Bossert, W. J. *Gen. Physiol.* **50**, 2061–2083 (1967).
4. Barry, P. & Hope, A. *Biophys. J.* **9**, 700–728 (1969).

THERMODYNAMICS

Nanoparticles behaving oddly

Steven G. Louie

CLUSTERS, or nanoparticles, are nanometre-scale aggregates of atoms. Their size gives them unusual optical, electronic and magnetic properties, caused by quantum-mechanical and Coulomb-charging effects, and they have many potential applications (see papers in a special issue of *Science*¹). As first predicted by Kubo² and others³ in the 1960s, metal nanoparticles at low temperatures show very different thermodynamic behaviour depending on whether they have an odd or even number of electrons, and on the fundamental symmetries of the electron Hamiltonian (that is, on how the energy and other properties of the electrons change under symmetry operations such as time reversal and space inversion). Some of these predictions have been confirmed in experi-

ments by Volokitin *et al.* (reported on page 621 of this issue⁴) on the electronic specific heat capacity and magnetic susceptibility of metal clusters.

For normal metals at low temperatures, the susceptibility (a measure of how easily a substance is magnetized) is independent of T , and the heat capacity is linear in T . These are both consequences of the Fermi–Dirac statistics of electrons having a finite and slowly varying density of states. But the electronic energy levels of a small cluster are discrete, because multiples of one-half of the electron wavelength must fit into the cluster. The average level spacing is inversely proportional to the number of electrons: the spacing in a bulk metal is infinitesimally small, but it can be several meV (corresponding to tens of

kelvin) for clusters containing only a few thousand valence electrons. At low temperatures, that discreteness will manifest itself in physical quantities.

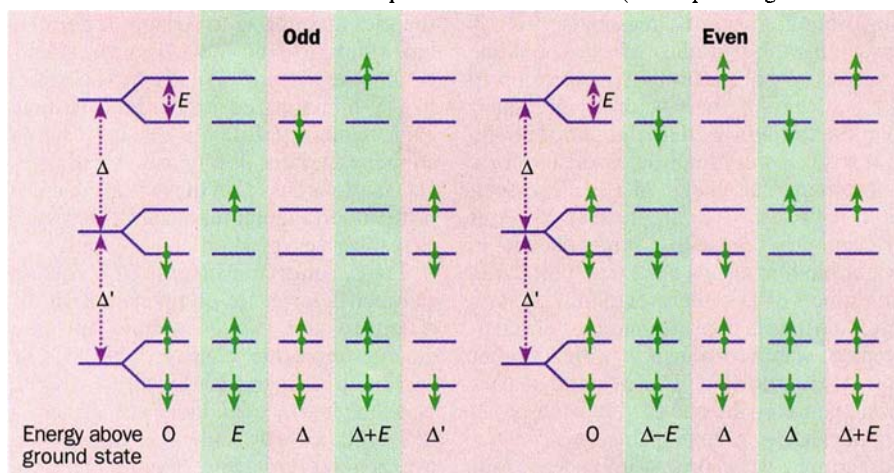
In practice, clusters of the same size may have different level spacings because of the extreme sensitivity of the electronic states to surface imperfections and shape. These irregularities cause admixtures between the quantum states of the ideal cluster, and change the energy levels. In quantum mechanics, these mixings or interactions are described by off-diagonal elements of the Hamiltonian matrix. So to understand the thermodynamic properties of clusters, a statistical treatment of the levels employing random matrix theory is used — the unknown interactions are modelled by random numbers. (Random matrices were first used in physics by Wigner⁵ to describe the distribution of observed resonances in nuclei. The idea has since found applications in atomic physics, mesoscopic systems, quantum chaos and particle physics.)

Because of quantum level-repulsion effects, however, the resulting distribution of levels is not one of uncorrelated random numbers but is dictated by the symmetries of the electron Hamiltonian, which depends on factors such as the strength of the spin–orbit interaction and the applied external magnetic field. The different level distributions lead to very different predicted low-temperature behaviours.

Moreover, because such small particles have a large charging energy, charge fluctuations will be suppressed at low T , and so the thermodynamics are determined by statistical averaging of systems each with a fixed number of electrons — that is, by using a canonical ensemble instead of a grand canonical ensemble (in which particle number is not conserved). That gives rise to different behaviours for clusters with odd or even numbers of electrons.

As illustrated in the figure, the low energy excitations of a cluster with an odd number of electrons correspond to those of a single free spin, leading to a susceptibility inversely proportional to temperature (the usual paramagnetic case). But clusters with an even number of electrons have a nondivergent susceptibility as $T \rightarrow 0$, because the spin-flip energy in this case is determined by the level spacing instead of the magnetic splitting energy.

Although quantum size effects have been seen in other properties of nanoclusters^{1,6,7}, their manifestations in the thermodynamic properties of metal nanoparticles had evaded detection so far. The major experimental problem was the lack of good samples, having a high density of isolated particles of uniform size and composition. With the advent of chemical synthesis of metal cluster compounds⁸, this problem has been overcome. The materials consist of identical clusters, each containing a core of closely packed metal atoms, capped with ligands. In their work,



Odd and even behaviour: energy-level diagrams for a metal nanoparticle with an odd number (left panel) and an even number (right panel) of electrons, in a magnetic field. The levels corresponding to opposite spins are usually degenerate, but the electron's energy depends on the alignment of its spin (equivalently its magnetic moment) relative to the field, giving rise to a splitting, E , proportional to the field. In each case the ground state is at the left, with successively higher-energy excited states to the right.

Volokitin *et al.*⁴ employed compounds with cores containing five, seven and eight shells of palladium atoms (561, 1,415 and 2,057 atoms, respectively). A palladium colloid with about 1.25×10^5 atoms per cluster was also studied.

Instead of being independent of temperature, the susceptibility increases sharply at a size-dependent onset temperature, and decreases again as T is lowered further. Although similar upturns had been seen and attributed to magnetic impurities, the size-controlled samples in this work allowed a quantitative identification of the upturn as a paramagnetic contribution from the odd-electron particles (the clusters have different and random surroundings, so about the same number are odd as even). The subsequent downturn can be explained by the large dipolar magnetic interactions between spins on neighbouring odd-electron particles, which lift the spin degeneracy of the electron levels, suppressing the susceptibility divergence.

The heat capacity of bulk metal is proportional to temperature. The clusters again showed a clear deviation from this: below one kelvin, the 5-shell heat capacity went through a transition to a T^2 dependence. This is the first observation of quantum size effects in the heat capacity of metal particles. The T^2 behaviour is expected of palladium clusters, which have weak spin-orbit interactions. The measured magnetic-field dependence further confirms the presence of unpaired spins on odd-electron particles.

Volokitin *et al.* make a compelling case for the presence of quantum size effects in the thermodynamic properties of metal nanoparticles. Their results provide the first experimental glimpse into the effects of having odd and even numbers of electrons, level repulsion and canonical ensemble statistics on the behaviour of metal clusters. Studies of this kind are vital to our fundamental understanding and practical exploitation of these systems. Our increasing ability to make clusters of varying types and sizes, together with other recently discovered structures such as quantum dots, fullerenes, nanotubes and nanocrystals, is opening up an exciting new arena for fundamental and applied research on the nanometre scale. □

Steven G. Louie is in the Department of Physics, University of California at Berkeley, and the Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

Same clock, different works

Paolo Sassone-Corsi

CIRCADIAN (from the Latin *circa-dies*) rhythms control most biological systems, from daily oscillations in plant photosynthesis to hormone secretion and annual breeding cycles in mammals. They have been described in nearly every eukaryote — including plants, fungi, insects and mammals — as well as in several prokaryotes. Circadian rhythmicity follows external geophysical cues, in particular the day–night cycle. Yet self-sustained rhythmic behaviour is still observed under constant environmental conditions, and this is a property of the so-called endogenous clock or pacemaker.

The first clock gene to be isolated was the *period* (*per*) gene from *Drosophila melanogaster*¹, which is essential for rhythmic adult eclosion (emergence from the pupa) and locomotor activity. PER is thought to represent a model for clock proteins², and two reports in the November issue of *Neuron* by Steven Reppert and colleagues^{3,4} shed new light on how it may act. The papers also show that expression of the *per* gene from the giant silkworm *Antheraea pernyi* is controlled in a completely different way, and that the PER protein seems to have hitherto unknown regulatory properties (see table).

It is generally accepted that pace-makers help to anticipate the needs of an organism through the cyclic regulation of specific target genes. Studies in various organisms have established that the endogenous clock consists of molecular cogs that reside and operate within individual cells. Attempts have been made to isolate clock genes in various organisms; however, as a clock can both measure and show time, it is necessary to distinguish the genes that are directly implicated in the clockwork from those that mediate output signals.

In the fruitfly, PER protein is thought to regulate its own expression through a negative-feedback loop⁵. Although it is still not clear whether the PER feedback loop involves direct transcriptional autoregulation, such loops seem to be a common feature of many clock-regulated systems⁶, such as the *frequency* gene of *Neurospora crassa*⁷ and the *CREM* gene in mammals⁸. The molecular elements of these feedback loops differ structurally between various systems, but the basic regulatory features do appear to be similar.

Expression of the *Drosophila* gene *timeless* (*tim*) is also controlled by an autoregulatory feedback loop, and its product,

TIM, forms dimers with PER^{2,9}. The PER and TIM autoregulatory loops are interdependent, and PER and TIM also rely on each other for translocation into the nucleus¹⁰. So timing of nuclear entry of the PER–TIM heterodimer in response to the day–night cycle is thought to be a key element that underlies the circadian clock in the fruitfly.

Because circadian clocks are common to most living organisms, and their fundamental properties seem to be highly conserved, it is reasonable to search for clock genes that could be conserved between species. But despite attempts made in several labo-

Characteristics of the *per-tim* system in the fruitfly and silkworm

	<i>Drosophila melanogaster</i>	<i>Antheraea pernyi</i>
Number of cells expressing PER in the brain	Dozens of neurons and > 100 glial cells	8 neurons
PER localization	Nuclear	Cytoplasmic
TIM localization	Nuclear	Cytoplasmic
<i>per</i> regulation	Regulatory feedback loop	Dynamic hybridization with antisense RNA
PER translation	Delayed	Synchronous with transcription

ratories, until last year *per* homologues had not been identified in any species other than dipterans (two-winged insects). This changed with the cloning of the *per* gene in the giant silkworm *A. pernyi*, an insect that diverged from dipterans 240 million years ago. These studies showed that the silkworm *per* gene is expressed in an oscillatory manner and that it can function as a circadian-clock element when substituted for the fruitfly *per* gene in arrhythmic *per*⁰ fly mutants¹¹. Moreover, in the silkworm, *per* seems to regulate the rhythmically gated process of egg hatching⁴, as has been observed for *Drosophila*¹. Along with the fact that key structural features of the *per* gene are conserved between the fruitfly and the silkworm, all of these results predicted that the mechanisms by which PER modulates the circadian clock in the two species may be the same. But the new work^{3,4} indicates that there are dramatic differences, which demonstrate that the same clock gene is able to act by two completely different mechanisms.

Sauman and Reppert³ have shown that only eight neurosecretory cells express the PER protein in the adult silkworm brain, in contrast to the dozens of neurons and hundreds of glia that express it in the fruitfly. Levels of PER oscillate in these neurosecretory cells, and levels of TIM protein also cycle in a coordinate manner here. Surprisingly, PER is never found in the nucleus — immunostaining using specific antibodies shows a characteristic

1. Science **271**, 920–945 (1996).

2. Kubo, R. J. *Phys. Soc. Jap.* **17**, 975–986 (1962).

3. Halperin, W. P. *Rev. Mod. Phys.* **58**, 553–606 (1986).

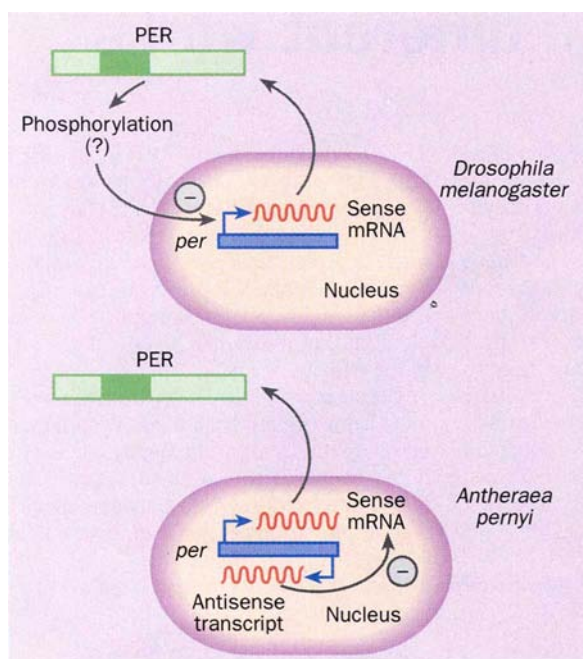
4. Volokitin, Y. *et al.* *Nature* **384**, 621–623 (1996).

5. Wigner, E. P. *Proc. Camb. Phil. Soc.* **47**, 790–798 (1951).

6. Cohen, M. L. & Knight, W. D. *Physics Today* **43**, 42–50 (1990).

7. Kastner, M. A. *Physics Today* **46**, 24–31 (1993).

8. de Jongh, J. J. (ed.) *Physics and Chemistry of Metal Cluster Compounds* (Kluwer, New York, 1994).



Two distinct mechanisms regulate the circadian clock in the fruitfly and in the silkworm. Reppert and colleagues^{3,4} have shown that the protagonist is the same, but the way in which it acts is different. Whereas in the fruitfly, *per* expression is regulated by a negative-feedback loop, in the silkworm, regulation of the *per* mRNA transcript is achieved by dynamic hybridization with antisense RNA.

'doughnut', which indicates that the protein does not leave the cytoplasm. Levels of PER also oscillate in the axonal projections that originate from the eight neurosecretory cells, and PER levels correlate with the cyclic output of neuropeptide secretion. So the clock function of the PER-TIM dimer must operate without entering the nucleus and, therefore, independently of transcriptional autoregulatory loops. Then the question is: how does the *per* transcript oscillate without the transcriptional feedback loop?

The surprise came from *in situ* hybridization studies. Analysis with the *per* sense strand, which is classically used as an experimental control, showed a bright signal that also oscillated with a circadian periodicity. Although the exact structure of the silkworm *per* gene is not yet known, Reppert and colleagues believe that an antisense *per* transcript may be generated by an intronic promoter (see figure). The antisense transcript can hybridize with the sense messenger RNA and block its function. In this way, the oscillating levels of PER protein can be determined.

But how can the oscillating PER modulate levels of TIM in the absence of a nuclear feedback loop? This is still not clear, although it could be argued that *tim* transcripts are regulated in a similar fashion. It has been proposed that, in the fruitfly, the PER-TIM dimer acts directly on the transcription of target genes, resulting in the regulation of output rhythms². This is clearly not the case in silkworm neurons, because the two proteins do not enter the

nucleus. So if the PER-TIM complex is truly a key clock component, there must be alternative pathways for the regulation of overt rhythms.

Notably, the cytoplasmic localization of PER in the silkworm is seen only in the eight neurosecretory cells in the brain — in other silkworm cells, such as in the embryonic gut epithelium, PER is, in fact, nuclear, and functions in the same way as the fly protein. Moreover, translation of PER protein in silkworm brain cells is synchronous with transcription of the *per* gene. So the transcription-to-translation event provides no delay in this circadian timing system. This is in contrast to the delay in translation of four to six hours that has been observed in the fruitfly^{5,12}. On the other hand, the pattern of *per* transcription-translation in the silkworm eye shows the same temporal delay as transcription-translation of the *Drosophila*

per gene. So two completely different regulatory pathways are acting in the same organism and on the same gene — in ocular photoreceptors and neurons — to generate an apparent equivalent rhythmicity.

The challenge is to understand the molecular features of the clock. The finding by Reppert and colleagues that *per* oscillation in the *A. pernyi* brain is due to a dynamic synthesis and pairing of sense and antisense transcripts is a tantalizing new twist to the story. The relative half-lives of these mRNAs may be a key determinant of the circadian behaviour of the silkworm, and experiments aimed at modifying this behaviour should prove to be exciting. □

Paolo Sassone-Corsi is at the Institut de Génétique et de Biologie Moléculaire et Cellulaire, BP 163, 67404 Illkirch, Strasbourg, France (e-mail: paolosc@titus.u-strasbg.fr).

1. Konopka, R. J. & Benzer, S. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
2. Hall, J. C. *Trends Neurosci.* **18**, 230–240 (1995).
3. Sauman, I. & Reppert, S. M. *Neuron* **17**, 889–900 (1996).
4. Sauman, I., Tsai, T., Roca, A. L. & Reppert, S. M. *Neuron* **17**, 901–909 (1996).
5. Hardin, P. E., Hall, J. C. & Rosbash, M. *Proc. Natl Acad. Sci. USA* **89**, 11711–11715 (1992).
6. Sassone-Corsi, P. *Cell* **78**, 361–364 (1994).
7. Aaronson, B. D., Johnson, K. A., Loros, J. J. & Dunlap, J. D. *Science* **263**, 1578–1584 (1994).
8. Molina, C. A., Foulkes, N. S., Lalli, E. & Sassone-Corsi, P. *Cell* **75**, 875–886 (1993).
9. Myers, M. P., Wager-Smith, K., Wesley, C. S., Young, M. W. & Seghal, A. *Science* **270**, 805–808 (1995).
10. Seghal, A. et al. *Science* **207**, 808–810 (1995).
11. Levine, J. D., Sauman, I., Imbalzano, M., Reppert, S. M. & Jackson, F. R. *Neuron* **15**, 147–157 (1995).
12. Curin, K. D., Huang, Z. J. & Rosbash, M. *Neuron* **14**, 365–372 (1995).

DAEDALUS

Winter's icy grip

ICE and snow are very slippery, a fact exploited to the full in skating and skiing, but a major hazard on ice-bound roads. One explanation is that ice contracts on melting. Its melting-point therefore decreases with pressure. Press on it with a skate or a tyre, and it finds itself above its melting-point. It promptly melts to a thin film of lubricating water. Such light pressure may seem inadequate; but only a tiny fraction of a solid surface can ever make true molecular contact with another, so the local pressures are enormous.

Musing on this problem, Daedalus began to invent a refrigerated tyre. It would be so much colder than ambient that even the highest pressure could not melt the ice to water. But he soon realized that if the tyre happened to hit a patch of wet ice, or a puddle of slush, it would solidify the water instantly and freeze in place. The vehicle would jerk to an immovable halt.

He then recalled the intriguing fact that stretched rubber cools when it is allowed to contract, and warms up again when stretched. A normal tyre is too inextensible to exploit this effect. What is needed is a pneumatic tyre that stretches markedly on being inflated. Where it hits the road, it will be compressed by the weight of the vehicle, and will cool down. On a wet road it will freeze in place, secure against all skidding. Yet the stretching of the trailing edge as it lifts away from the road will warm it up again, and release the grip. This cunning tyre will freeze firmly to a wet road, and yet roll freely. Sadly, practical tyre rubbers may not cool strongly enough for the job.

So Daedalus is turning his attention to the wet road surface itself. What is needed, he says, is a tyre coating that under pressure reacts reversibly with water or snow, forming a solid hydrate. Many such hydrates are known, mainly of small molecules such as bromine or butane. He is now devising tyre-coatings loaded with hydratable groupings. When the 'Hydratyre' hits a wet or icy road, the huge local pressure will form solid hydrate, binding it to the road. As it lifts away, the pressure will be released, and the hydrate will decompose again. In wet or frost, the Hydratyre will be utterly skid-proof. In the dry, however, its non-rubbery coating may be a trifle slippery.

This paradoxical property may pose novel hazards. Amazed at how surely they can drive in wet and snow, drivers may be tempted into still more frightful risks on dry roads. Until all drivers have adjusted to their new Hydratires, road authorities may have to minimize the danger by spraying water or ice on the roads in hot weather.

David Jones

The birds and the bees

SIR — Many Lorantheae mistletoes, including *Peraxilla* in New Zealand, have 'explosive' flowers which cannot open themselves; birds pop the buds open and pollinate them¹. However, a reduction in bird density has decreased visitation rates in some areas and depressed seed production in these plants². Here we report that tiny native solitary bees (*Hylaeus* sp.) can successfully prise open *Peraxilla tetrapetala* buds, which is to our knowledge the first documented case of an invertebrate opening an explosive, bird-pollinated flower. The bee is inefficient at pollen transfer, but still doubles the number of seeds ripened per flower. This example is an extreme case of ecological replacement of vertebrates by an invertebrate in a depauperate island fauna.

Peraxilla flowers are relatively large (27–42 mm); ripe buds open only with considerable force, from a beak or a human hand, for example. The main pollinators are honeyeaters¹: tui (*Prothemodera novaeseelandiae*, mass 120 g) and bellbirds (*Anthornis melanura*, 30 g). Explosive flowers worldwide are usually bird-opened, exclusively so in the Lorantheae³. Insect-tripped explosive flowers occur in several families, such as *Cytisus scoparius* (Fabaceae)^{4,5}, but there are no previous reports of any explosive flower that can be

opened by both birds and insects.

We have observed native solitary bees (*Hylaeus* sp.; Hymenoptera, Colletidae) opening flowers of *P. tetrapetala* at two sites, Craigieburn and Lake Ohau. The small bees (body length 7 mm) are dwarfed by mistletoe buds (see figure), yet at Lake Ohau bees successfully opened about one bud in four by attacking the tip of the bud with their mandibles. This takes considerable effort and time (20–40 seconds), whereas tui open buds in 0.23 ± 0.06 seconds¹. As *Hylaeus* weigh only 0.01 g (1/3,000 of the bellbird mass) it is remarkable that they can open buds at all. Insect-explosive flowers like *Cytisus scoparius* are much more restrictive; bumblebees trip them easily, whereas slightly lighter honeybees have difficulty⁶.

Peraxilla stigmas are well placed to collect pollen from an approaching bird, but *Hylaeus* are pollen harvesters⁷ and are so small that they often do not touch the stigma, reducing pollination effectiveness. Our experiments on plants in the absence of birds show that bees cannot open *P. colensoi* flower buds, even though they visit flowers frequently. *P. colensoi* buds require more force to open than *P. tetrapetala*, and may be beyond the strength of these bees.

In contrast, bees open many *P. tetrapetala* buds inside our experimental wire-mesh bird-exlosures; significantly more of these opened flowers ripen a seed (see table). A few unopened buds set seeds by self-pollination in both *Peraxilla* species¹, but by opening *P. tetrapetala* buds *Hylaeus* doubles the fruit-set. As *Hylaeus* cannot readily contact the stigma, it could improve the fruit-set partly by giving other invertebrates access to open flowers. Introduced wasps (*Vespula* spp.) and honeybees (*Apis mellifera*) were present at all sites that we studied, but do not open *Peraxilla* flowers. These flowers are opened only by native animals (whether birds or insects).

New Zealand generally has unspecialized pollination syndromes⁸ and a depauperate fauna (there are no land mammals). This could explain why the only invertebrate known to open bird-adapted explosive flowers is found in New Zealand. Elsewhere, changes in pollinator availability in isolated



A native bee (*Hylaeus* sp.) harvesting pollen from the explosive-flowered mistletoe *P. colensoi* at Wakefield, New Zealand. Although the flowers are normally bird-opened, *Hylaeus* can open flowers of *P. tetrapetala*, but not *P. colensoi*. The bee is so small that it may not regularly contact the stigma in either species, but *P. tetrapetala* flowers opened by these bees do ripen more seeds.

populations may promote speciation via pollination syndrome shifts, like the one recently analysed genetically in *Mimulus*⁹. The tiny *Hylaeus* is an unlikely substitute for birds on large flowers like *P. tetrapetala*, yet we have found that in the absence of birds it can double seed production. Because habitat clearance and introduced mammalian predators in New Zealand have lowered both bird¹⁰ and mistletoe¹¹ densities, this small pollen harvester provides important conservation benefits.

Dave Kelly

Jenny J. Ladley

Plant and Microbial Sciences,
University of Canterbury,
Christchurch 1, New Zealand
e-mail: D.Kelly@botn.canterbury.ac.nz

Alastair W. Robertson

Ecology Department, Massey University,
Palmerston North, New Zealand

Joan Edwards

David C. Smith

Biology Department, Williams College,
Williamstown, Massachusetts 01267, USA

Fruit-set in *Peraxilla* under various pollination treatments

Pollination treatment	<i>P. tetrapetala</i>	<i>P. colensoi</i>
Unopened	16.3*	11.0*
Bee-opened, birds excluded	30.7†	not applicable
Hand-opened, birds excluded	48.2††	15.8†
Control, naturally opened	28.3**	35.4†
Hand-opened, hand-pollinated	66.3‡	not done
Residual variance (with residual d.f.)		
Model: Null	176.8 (34)	177.8 (17)
Plant	88.55 (27)	107.2 (12)
Plant, treatment	31.44 (23)	33.7 (10)
Probability (treatment)	< 0.001	< 0.001

Fruit-set is defined as the percentage of flowers ripening their solitary ovule. We carried out treatments in summer 1995–96 in the field at Craigieburn, Canterbury (43° 9.2' S, 171° 42.7' E) for *P. tetrapetala* and at Wakefield, Nelson (41° 25.5' S, 173° 2.5' E) for *P. colensoi*. Pollination treatments were as follows. Unopened: buds within wire-mesh bird-exlosures which remained unopened until petal abscission, see ref. 1; bee-opened, birds excluded: opened by bees within the enclosures; hand-opened, birds excluded: buds were opened with human pollen transfer minimized and left accessible to insects; controls: flowers accessible to birds and insects and naturally opened; hand-opened hand-pollinated: abundant outcross pollen but birds excluded. All treatments were replicated within plants as blocks ($n=6$ for *P. colensoi* and 7 for *P. tetrapetala*) and analysed for each species using a generalized linear model with binomial error and logit link function. In each species treatment was highly significant. Comparisons among means were done using *t*-tests on generalized linear model-generated means (and s.e.m.) for pairs of treatment groups; means within a species indicated by the same symbols (*†‡) do not differ significantly ($P>0.05$).

1. Ladley, J. J. & Kelly, D. *Nature* **378**, 766 (1995).
2. Ladley, J. J. & Kelly, D. *For. Bird* **278**, 16–21 (1995).
3. Kuijt, J. *The Biology of Parasitic Flowering Plants* (Univ. California Press, Los Angeles, 1969).
4. Proctor, M. C. F. & Yeo, P. F. *The Pollination of Flowers* 201 (Collins, London, 1973).
5. Aluri, R. J. S. & Subba Reddi, C. *Proc. Indian Natl. Sci. Acad. B* **61**, 323–332 (1995).
6. Meeuse, J. D. *The Story of Pollination* (Ronald, New York, 1961).
7. Donovan, B. J. & Macfarlane, R. P. in *New Zealand Pests and Beneficial Insects* (ed. Scott, R. R.) 247–264 (Lincoln Univ. Coll. Agriculture, Canterbury, 1984).
8. Webb, C. J. & Kelly, D. *Trends Ecol. Evol.* **8**, 442–447 (1993).
9. Bradshaw, H. D. Jr, Wilbert, S. M., Otto, K. G. & Schamske, D. W. *Nature* **376**, 762–765 (1995).
10. King, C. M. (ed.) *The Handbook of New Zealand Mammals* (Oxford Univ. Press, Oxford, 1990).
11. Ogle, C. & Wilson, P. *For. Bird* **237**, 10–13 (1985).

Chiral sculptured thin films

SIR — We have observed optical activity in thin solid films composed of nanometre-scale helical columns. These films rotate the plane of polarization of light in a manner analogous to cholesteric liquid crystals and other structurally chiral media. This rotation demonstrates the optical activity of the helical structured films and indicates that these films possess other significant optical properties such as circular dichroism (the effect currently exploited in cholesteric liquid-crystal displays). Our measurements demonstrate the value of a new thin-film deposition technique^{1,2} called glancing angle deposition (GLAD), which allows three-dimensional control of film structure on a scale of less than 10 nm.

In structurally chiral media, a helical structure creates optical activity. For example, in cholesteric liquid crystals (CLCs), the helical structure of the stacked layers of molecules alters the vibration ellipse of plane-polarized light³. Electromagnetic properties of chiral media have been investigated with light in

an angle θ of 85° measured from the substrate normal. Under conditions of almost grazing incidence ($\theta > 75^\circ$), local film-growth variations lead to strong shadowing effects and competition between nucleating grains. The high points of the film intercept most of the flux, and the low points become shadowed and do not grow further. This leads to a porous film with isolated columns of material inclined towards the source of the flux.

We have taken advantage of this behaviour and incorporated a rotation of the substrate about an axis normal to the surface, to 'sculpt' the columns into helices, as shown in Fig. 1. Using feedback from a deposition rate monitor, the rotation speed of the substrate can be controlled to produce helices of a predetermined pitch. We have fabricated helices of similar form with pitches ranging from 50 to 2,000 nm, and in films composed of MgF_2 , SiO_2 , CaF_2 , chromium, manganese and copper.

When linearly polarized light is incident on a helical sculptured thin film along the film's helical axis, the transmitted light exhibits a rotation of the plane of polarization. This optical rotation is a function of the incident light wavelength in a vacuum λ_{vac} , and the phenomenon is called optical rotatory dispersion⁵. The optical rotatory dispersion of the 17.3-turn MgF_2 helical thin film (Fig. 1) was determined for a range of wavelengths, $560 \text{ nm} \leq \lambda_{\text{vac}} \leq 633 \text{ nm}$, in the visible spectrum (Fig. 2). From comparison with CLCs, we expect that optical rotation will be largest when the pitch (P) equals some 'average' wavelength of light in the film λ_{vac} divided by a suitably averaged refractive index n_{avg} of the film. To estimate n_{avg} of the MgF_2 chiral thin film, we used density measurements and a simple mixing rule⁶. With $n_{\text{avg}} = 1.1$ and $P = 360 \text{ nm}$, we expect the MgF_2 thin film to show very large optical rotation when $\lambda_{\text{vac}} \approx 400 \text{ nm}$. In the same regime, we expect strong circular dichroism (depolarization of linearly polarized light).

We measured optical rotations for enantiomeric pairs of sculptured thin films (an identical pair where one has exclusively left-handed helices and the other has exclusively right-handed

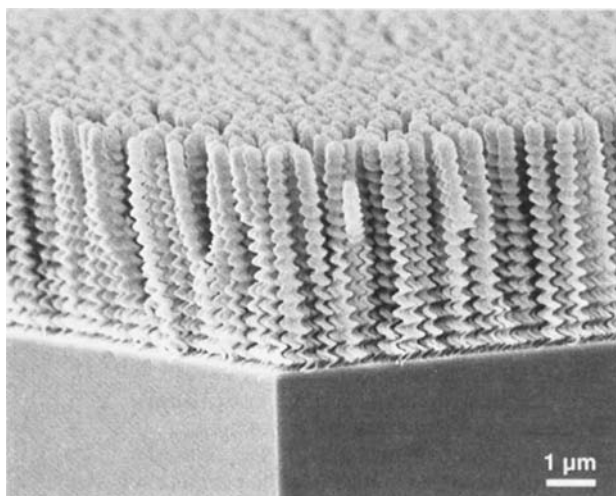


FIG. 1 Scanning electron micrograph of a magnesium fluoride sculptured thin film on a glass substrate. The film has 17.3 helical turns of pitch 360 nm.

CLCs, with microwaves in wire spirals⁴ and with suitable wavelength radiation in other helically structured materials. We have previously described the use of a new fabrication technique to produce helical electromagnetic properties in thin solid films of inorganic compounds. These 'sculptured' thin films are characterized by helical columns with pitches comparable to the wavelengths of visible light³.

We fabricated the films with a deposition system using oblique incidence flux and substrate motion (GLAD)^{1,2}. We evaporated source material in a vacuum chamber with a pressure of 1×10^{-6} torr ($\sim 1 \times 10^{14} \text{ Pa}$). By tilting the substrate, the evaporant flux was incident at

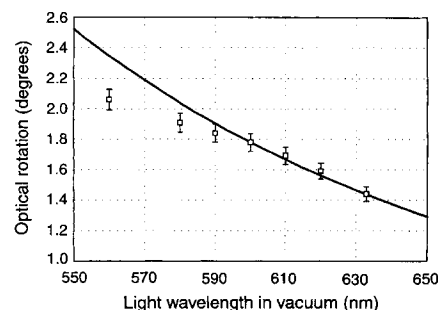


FIG. 2 Optical rotation as a function of wavelength (optical rotatory dispersion) of the magnesium fluoride chiral thin film shown in Fig. 1. Maximum optical rotation is expected when $\lambda_{\text{vac}} \approx 400 \text{ nm}$. The solid line shows the $\lambda_{\text{vac}}^{-4}$ behaviour predicted for CLCs by the de Vries formula at high values of λ_{vac} .

helices). We found measured values to be anticlockwise for the right-handed member and clockwise for the left-handed member of the enantiomeric pair, the magnitudes of the two optical rotations being identical. In addition, we observed no optical rotation for a glass substrate without a film.

Measurements of optical rotatory dispersion of helical sculptured thin films demonstrate their optical activity. With the widespread use of organic chiral materials, these artificially produced chiral structures of inorganic materials seem to be a promising new optically active medium.

K. Robbie

M. J. Brett

Department of Electrical Engineering and
Alberta Microelectronic Centre,
University of Alberta,
Edmonton, Alberta T6G 2G7,
Canada

e-mail: robbie@ee.ualberta.ca

A. Lakhtakia

Department of Engineering Science and
Mechanics,
Pennsylvania State University,
University Park,
Pennsylvania 16802-1401, USA

- Robbie, K., Brett, M. J. & Lakhtakia, A. *J. Vac. Sci. Technol. A* **13**, 2991–2993 (1995).
- Lakhtakia, A., Messier, R., Brett, M. J. & Robbie, K. *Innov. Mater. Res.* **1**, 165–176 (1996).
- de Gennes, P. G. & Prost, J. *The Physics of Liquid Crystals* (Clarendon, Oxford, 1993).
- Lindman, K. F. *Ann. Phys. Lpz.* **63**, 621–644 (1920).
- Barron, L. D. *Molecular Light Scattering and Optical Activity* (Cambridge Univ. Press, Cambridge, 1982).
- Lakhtakia, A. (ed.) *Selected Papers on Linear Optical Composite Materials* (SPIE Optical Eng., Bellingham, 1996).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words and ten references.

Fault lines in verse

David R. Oldroyd

Wordsworth and the Geologists. By John Wyatt. Cambridge University Press: 1996. Pp. 268. £35.

WORDSWORTH'S life and work have been combed by regiments of scholars. But perhaps there was a corner that had not quite received its due: Wordsworth's interest in geology, and his connections with geologists. Now John Wyatt scrutinizes this corner in minute detail. He describes the state of geology in Wordsworth's day, gives an account of Wordsworth's geological knowledge, describes the network of 'Trinity men' associated with Wordsworth and argues that geology, in Wordsworth's day, was a "poetic discipline".

Investigations into the connections between art and science sometimes clutch at straws — at any poetic allusions that offer even a hint of 'influence' between science and literature. Wyatt resists this tendency, rightly indicating where claims that might be advanced by a less prudent scholar cannot be sustained. But, as a result, he has to admit that his examples of common content in the writings of Wordsworth and his scientific contemporaries are few. So, following Rom Harré, he looks for commonalities of style and shared high-level general concepts such as "decay and renewal" or "universality".

Not surprisingly, such general similarities can be found readily enough. But, having found them, not all will concede that they are profoundly significant; so I leave it to others to decide whether they want to get excited about the proposition that both natural philosophers and men of letters at the beginning of the nineteenth century were concerned about, say, decay and renewal.

Harré was interested in the ocean of conversation in which we are immersed, and saw it as the 'reality' that binds humans into a society. So it behoves Wyatt to examine the conversations Wordsworth may have had, or at least the persons with whom he may have conversed on scientific subjects; and this leads him to an examination of Wordsworth's circle of friends. It included the Trinity men: Wordsworth's brother, Christopher, master of Trinity; Adam Sedgwick, the great dalesman geologist; and the Lancastrian William Whewell, master of everything, including mineralogy and geology.

Starting from these men, Wyatt works outwards, examining an important part of Wordsworth's large social network.

As is well known, Wordsworth wrote a successful *Lakeland Guide*, later editions of which included "letters" by Sedgwick on the geology of the region. Wyatt points out that successive editions of the *Guide* reveal growing interest in matters geological. That is true. But the growth started from a low base, and nowhere did Wordsworth offer anything remotely resembling a scientific treatment. Nor, I believe, was he in any position to do so. That was left to Sedgwick.

What was the relationship between the two friends? Sedgwick undoubtedly visited Wordsworth when in the Lakes. They dined together, and walked on the hills. Socially, they had much in common. Yet in a way they were worlds apart. In his pioneering Lakeland work, Sedgwick was forever measuring dips and strikes, hammering rocks and identifying them in an approximate fashion, noting outcrops and recording exposures of different rock types on maps. He sought to ascertain the laws governing the arrangement and formation of mountains. Ultimately, he sought to subsume his observations under the head of Élie de Beaumont's

theory of the Earth. In this, Sedgwick was following the Cambridge tradition of a mathematical approach to science, properly remarked on by Wyatt.

But it is doubtful whether Wordsworth understood, or sympathized with, Sedgwick's task. In *The Excursion* we read:

He who with pocket hammer smites the edge
Of luckless rock or prominent stone...
...detaching by the stroke
A chip or splinter — to resolve his doubts;
And, with that ready answer satisfied,
The substance classes by some barbarous
name,
And hurries on; . . .
— and thinks himself enriched,
Wealthier, and doubtless wiser than before.

Although this poem was published in 1814, the lines could describe Sedgwick rather well. Wyatt resists this suggestion, however, arguing that the words would not apply to Sedgwick. They refer, he says, to some mineral collector, not a geologist proper. Be that as it may, I believe that Wordsworth did not, in 1814 at least, think that the way to understand nature was to collect chips of rock and enter information on maps. (Admittedly, his views may have changed somewhat by the time he walked with Sedgwick in the 1820s.)

Nevertheless, Wyatt tries to show how Wordsworth was versed in geology, and that some geological understanding lies below the surface of his poetry. The *Dudon Sonnets*, for example, are represented as a "systematic description of a



ART meets geology in this drawing from 1817 of Fingal's Cave on the island of Staffa in the Inner Hebrides, Scotland. The artist, William Daniell, was striving for a more realistic image than that of his earlier, more fanciful, pictures of the cave. From *Science and the Perception of Nature: British Landscape Art in the Late Eighteenth and Early Nineteenth Centuries* by Charlotte Klonk. Yale University Press, \$55, £40.

river". Systematic it may be, in that the poems describe the course of the river from its source to the sea, but the description cannot possibly be described as scientific, and would yield no information useful to a geologist. Wordsworth was assuredly aware of slow processes of change, as, for example, when lakes fill with silt. But the puzzle is, I suggest, that Wordsworth showed so little geological interest, given his scientific acquaintances and the fact that he lived in a geological Mecca. I find no evidence that he developed geological theories or made systematic observations of rocks, strata, minerals or landforms.

Obviously, he loved the countryside and wrote with the pen of genius. But science was not his *métier*. He didn't attend meetings of the British Association, for example, where he could have consorted with scientists to his heart's content.

In point of fact, Sedgwick really wanted to see his letters published in the *Lakeland Guide* of his dear friend Jonathan Otley, Keswick clock repairer and amateur naturalist. But they had been promised to Wordsworth — to Sedgwick's regret. The humble Otley was a real naturalist. He guided Sedgwick to important exposures, taught him to distinguish bedding, cleavage and jointing, and proposed a threefold subdivision of the Lakeland rocks that Sedgwick adopted (and which stands to this day). In fact, judging by his field notebooks, Sedgwick spent more time in the Lakes with men like Otley, or the engineer and surveyor Joseph Fryer, than with social equivalents like Wordsworth.

What Wyatt does show successfully is Wordsworth's interest in metaphysical problems that also concerned contemporary scientists: natural causes, rates and processes of change, origins and ends, uniformity and progression, and so on. Sedgwick certainly thought about such matters. But they were subordinate to his scientific practice, and to his titanic struggles with fellow geologists, especially Murchison.

To be sure, Wyatt admirably displays the interconnections, social and mental, between the members of the community in which Sedgwick and Wordsworth moved. But, I believe, the divide between the two cultures was already emerging in the Romantic era, and attempts to reconcile them by detailed historical research, hunting for minuscule clues indicating "influences", or grand commonalities of a metaphysical nature, may not help our understanding of a largely inevitable process, already well advanced even in the early nineteenth century. □

David R. Oldroyd is at the School of Science and Technology Studies, University of New South Wales, Sydney 2052, Australia.

The beginning of the end

Harriet Coles

Sex and the Origins of Death. By William R. Clark. Oxford University Press: 1996. Pp. 224. £16.99, \$22. UK publication date is 30 January.

SOME organisms can survive extreme conditions in a state of suspended life — or reversible death — called cryptobiosis. Encysted embryos of the brine shrimp *Artemia salina* kept only 2.2 °C above absolute zero for six days hatch as successfully as those kept at room temperature.

The nineteenth-century scientists who saw cryptobiosis as a possible biological basis for resurrection are forgiven by William R. Clark for being "sophomoric" (*Oxford English Dictionary*: "pretentious, bombastic, immature, crude and superficial"). This is a precautionary measure by the author, who asks, in the light of late-twentieth-century evolutionary and cell biology, "what is death?"

Death is variously defined by Clark as the destruction of a genome, the loss of whole-brain function, the end-point of senescence, the genetically encoded self-deconstruction of cells (known as apoptosis), the end of a 'self' and (with due respect to Hinduism) what happens to a mouse when an elephant sits on it. In exploring each of these forms of death, Clark moves seamlessly from the rare genetic Hutchison-Gilford syndrome in which 13-year-old children die of old age, to the *p53* gene as a master switch between the life and death of a cell, and the ethical dilemma of maintaining more and more people in a persistent vegetative state.

The "sex" in the title of the book is more than a marketing ploy. Clark translates the age-old conceit of sex and death (which can be traced from the poetry of John Donne to Hollywood screenplays by way of Freud) into the language of evolutionary biology. One of the earliest sexual reproducers to evolve — the single-celled organism *Paramecium* — also evolved the programmed destruction of DNA used for cell maintenance but not of DNA set aside for reproduction. Seen from this perspective, multicellular organisms are an elaboration of this theme of division of labour into somatic and germline DNA. Clark eloquently traces the ancestral roots of both programmed cell death and, more indirectly, death of the whole organism to the evolution of sex in *paramecia*.

At best, this book is an enjoyable and eminently readable collection of light essays on the several definitions of death

and life, but it aspires to be much more. It is the second popular book in two years by Clark (a professor of immunology at the University of California at Los Angeles), and in it he goes where the most respected investigators of apoptosis and senescence fear to tread. He tries to arrange this smorgasbord of deaths around seemingly profound statements about life such as "Death is not inextricably intertwined with the definition of life", "...in multicellular organisms death always begins with the death of single cells" and "If life is the interaction of structure with energy, then it follows that death at the level of a single cell must represent the loss of either structure or energy". However dazzling I found these brain-teasers, it was disappointing to realize I had been made giddy by nothing other than my own all-too-human desire for ultimate answers.

This absorbing collection of scientific 'death' curiosities was further undermined by a fictitious emergency-room scenario (complete with cardiac arrest, defibrillators, bleeping monitors and tense paramedics) that Clark uses to link myocardial cell death to existential oblivion. If you have to be 'profound', do it in moderation and leave out the hype. □

Harriet Coles is an assistant editor of *Nature*.

New in paperback

African Exodus: The Origins of Modern Humanity

by Chris Stringer and Robin McKie. Pimlico, £12.50. "The defence [of the 'Out of Africa' model] is skilful and enthusiastic. Topics that often appear obscure and impenetrable to the public are made crystal-clear", Jean-Jacques Hublin, *Nature* **381**, 658 (1996).

Models of My Life

by Herbert A. Simon. MIT Press, \$22.50, £18.95. "A mine of shrewd observations... the autobiography covers a period of astonishing intellectual ferment", H. Christopher Longuet-Higgins, *Nature* **350**, 531 (1991).

Power from Wind: A History of Windmill Technology

by Richard L. Hills. CUP, £18.95, \$29.95. "An important contribution to the history of technology", A. G. Keller, *Nature* **372**, 627 (1994).

Image and Brain: The Resolution of the Imagery Debate

by Stephen M. Kosslyn. MIT Press, \$27.50, £23.50. Reviewed by Zenon Pylyshyn in *Nature* **372**, 289 (1994).

Hidden Attraction: The Mystery and History of Magnetism

by Gerrit L. Verschuur. OUP, £14.95, £10.99. Reviewed by L. Pearce Williams in *Nature* **364**, 586 (1993).

Gravitational scattering as a possible origin for giant planets at small stellar distances

Stuart J. Weidenschilling* & Francesco Marzari†

* Planetary Science Institute/SJI, 620 North Sixth Avenue, Tucson, Arizona 85705, USA

† Dipartimento di Fisica, Università di Padova, Via Marzolo 8, I-35131 Padova, Italy

THE recent discoveries¹⁻⁴ of massive planetary companions orbiting several solar-type stars pose a conundrum. Conventional models^{5,6} for the formation of giant planets (such as Jupiter and Saturn) place such objects at distances of several astronomical units from the parent star, whereas all but one of the new objects are on orbits well inside 1 AU; these planets must therefore have originated at larger distances and subsequently migrated inwards. One suggested migration mechanism invokes tidal interactions between the planet and the evolving circumstellar disk⁷. Such a mechanism results in planets with small, essentially circular orbits, which appears to be the case for many of the new planets. But two of the objects have substantial orbital eccentricities, which are difficult to reconcile with a tidal-linkage model. Here we describe an alternative model for planetary migration that can account for these large orbital eccentricities. If a system of three or more giant planets form about a star, their orbits may become unstable as they gain mass by accreting gas from the circumstellar disk; subsequent gravitational encounters among these planets can eject one from the system while placing the others into highly eccentric orbits both closer and farther from the star.

The most generally accepted model for the origin of the giant planets, Jupiter and Saturn, in our own Solar System assumes that solid planetesimals accumulated to form cores having masses of the order of ten times Earth's mass. Each core was then able to initiate accretion of gas from the surrounding solar nebula; the mass of the accreted gas became great enough for its own gravity to continue the process until that region of the nebula was exhausted, or tidal torques formed a gap at the planet's orbit⁵. Formation of such a core is believed to have required condensation of water (and perhaps other ices) to provide enough mass to allow growth of the core before dissipation of the gas. High temperatures within a few AU of the Sun would preclude such an origin. Jupiter, at 5.2 AU, is widely believed to have formed near the boundary of ice condensation in the solar nebula⁸. In contrast, at present six planets are known with masses comparable to Jupiter and orbits well inside 1 AU; these are companions of the stars 51 Pegasi, τ Boötis, ν Andromedae, 55 Cancri, HD114762 and 70 Virginis. The first three all have semimajor axes $a \approx 0.05$ AU, the next has $a \approx 0.1$ AU, and the last two are at 0.4 and 0.47 AU. In addition, a planet of 47 Ursae Majoris is at a distance of 2.1 AU.

Lin *et al.*⁹ suggested that giant planets migrated inwards while tidally linked to an evolving circumstellar disk that was accreting onto the star. This mechanism implies that planets must have formed simultaneously with the star itself, but it is difficult to account for the formation of planets on such short timescales⁶. Tidal linkage to the nebula also implies that any surviving planets should have essentially circular orbits. This appears to be the case for most of the newly-discovered planets, but 70 Vir B and HD114762 B have substantial orbital eccentricities ($e \approx 0.38$ and 0.25, respectively). This circumstance has led to the suggestion^{9,10} that these objects are brown dwarfs that formed during the collapse of their pre-stellar clouds, rather than planets that formed within circumstellar disks.

We suggest an alternative origin for these bodies that is compatible with the more conventional model of formation of jovian planets. The accretion of their solid cores from planetesimals is stochastic, and should produce multiple embryonic cores with the potential to accrete gas if their masses become sufficiently large. There is no *a priori* reason why these cores should form with such wide separations that the orbits of the final planets would remain stable after accretion of gas increases their mass by an order of magnitude or more. The likely outcome in many cases would be that the first core to accrete gas outstrips its neighbours, and potential rivals are either accreted or ejected from the system by its perturbations. However, stochastic growth of the cores will yield some systems with multiple cores in orbits that are stable before they accrete gas, but too close to be permanently stable after they gain additional mass from the nebula.

In the classical three-body problem, two planets initially in circular orbits cannot make close approaches if their semimajor axes, a , differ by more than $2\sqrt{3}R_H$, where R_H is the so-called Hill radius, equal to $\bar{a}[(m_1 + m_2)/3M_*]^{1/3}$, where $\bar{a} = (a_1 + a_2)/2$ is

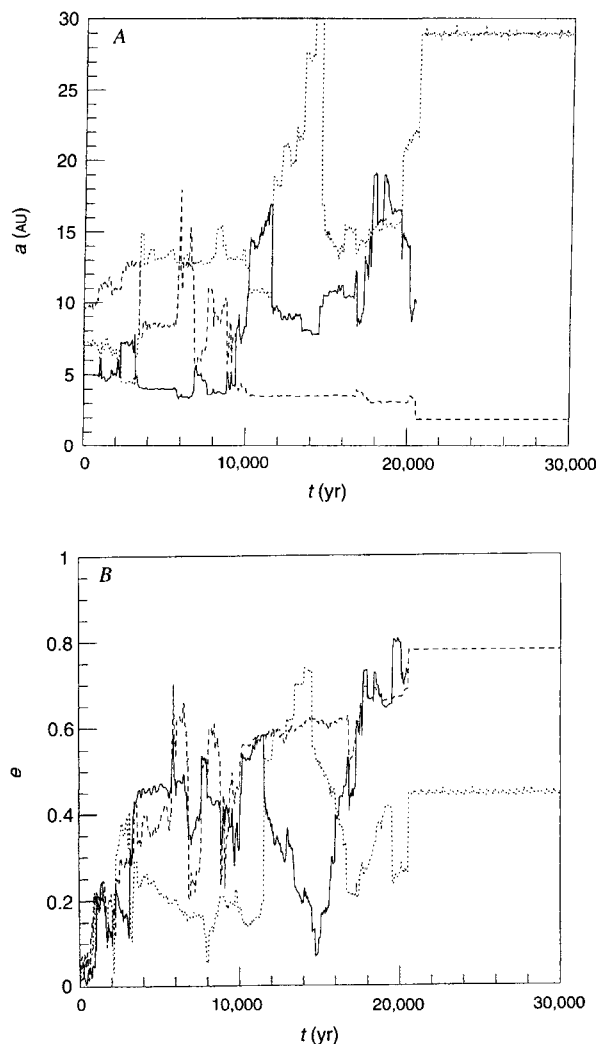


FIG. 1 Evolution of semimajor axes (panel A) and eccentricities (panel B) of a trio of Jupiter-mass planets orbiting a solar-mass star. Initial orbits are circular and coplanar, at distances of 5.0, 7.25 and 9.5 AU from the star. Their orbits become crossing, and after a series of close encounters the original inner planet is ejected hyperbolically at 21,000 yr. The planet originally in the outermost orbit is left in a close orbit ($a = 2$ AU, $e = 0.78$). The remaining planet has a distant orbit ($a = 29$ AU, $e = 0.44$).

their mean distance from the star of mass M_* , and m_1, m_2 are the planetary masses¹¹. If their orbital separation is less than this critical value, close encounters can occur. Gravitational scattering can change their orbits, but cannot produce stability; the ultimate fate of such a system is collision between the two bodies. Conservation of angular momentum decrees that the orbit of the combined body is between those of the original pair, and so does not extend the range of allowed orbits.

Very different outcomes are possible for a system of more than two planets. Chambers *et al.*¹² found that for three or more planets of comparable mass, mutual perturbations could produce crossing orbits, even though each pair of bodies considered separately meets the Hill stability criterion. There appears to be no absolute stability in such systems; the timescale for orbits to become crossing merely increases with initial separation. With three (or more) planets, a greater variety of outcomes is possible as the result of close encounters. Angular momentum can be partitioned unequally among the planets, and the system need not remain completely bound to the star. The idea that systems of massive planets could be dynamically unstable was first mentioned by Farinella¹³ in 1980. However, this suggestion was forgotten, as no such planets had been detected at that time, and there was no consensus as to how such systems could form.

We have carried out a series of numerical integrations of the orbits of three Jupiter-mass planets about a solar-mass star, using Everhart's¹⁴ integrator. The initial orbits were circular, and separated by 4–5 Hill radii. Simulations were performed for coplanar orbits and for small ($1\text{--}2^\circ$) initial inclinations, with similar outcomes. Mutual perturbations rapidly increased eccentricities until close approaches became possible. There followed a period of chaotic evolution with close encounters among the planets. The vast majority resulted in one planet being ejected on a hyperbolic trajectory. The remaining two planets usually had stable orbits. If the initial orbits were not strictly coplanar, the final orbits could also have significant mutual inclinations. There was no preference as to their fate; any of the three might be ejected, or end up in either the inner or outer orbit.

One example of such an outcome is shown in Fig. 1. Three Jupiter-mass planets had initially circular orbits at $a = 5.0, 7.25$ and 9.5 AU . After about 20,000 yr, the planet originally in the innermost orbit was ejected. The outermost original planet was left in a close orbit with $a = 2\text{ AU}$, $e = 0.78$, while the planet initially in the middle had a distant orbit with $a = 29\text{ AU}$, $e = 0.44$. These final orbits are stable against further encounters. The starting conditions were chosen to minimize computation time, and resulted in evolution on a timescale much shorter than the $\sim 10^6\text{ yr}$ needed for the planets to reach their assumed masses by accretion of gas⁶. We have carried out other simulations with larger initial separations, in which orbits became crossing only after several million years.

Although ejection of one planet is by far the most common outcome, other end states are possible. A small fraction of cases end with collisions between planets, or sequential ejection of two planets, leaving a single planet orbiting the star. Impact of a planet onto the star itself cannot be ruled out, though we have no examples of such an outcome. It is beyond the scope of this Letter to explore the full range of possibilities. We note that such systems are chaotic, that is, the outcome is sensitive to small differences in initial conditions. Even for a fixed set of planetary masses and initial semimajor axes, different angular separations at the start of the simulation can lead to very different final configurations. Much additional work will be needed to map out the full range of outcomes and their statistical distributions as functions of the masses and initial orbits of the planets. The present work established that dynamical interactions can alter the cosmogonically imposed initial state of a planetary system.

Wetherill^{15,16} has pointed out that there are problems with the timescale of formation of gas-giant planets by the core-accretion model (although these are less severe than for origin within a still-evolving disk). These considerations, and the lack

of detection of such planets in earlier surveys, led him to suggest that they might be very rare¹⁶. If so, then it is surprising that our own system contains two such bodies, Jupiter and Saturn. It seems plausible that other planetary systems could produce three (or perhaps more) gas giants, particularly if their circumstellar nebulae were more massive than our own. If systems of multiple giant planets in unstable orbits are a common outcome of stochastic core formation followed by gas accretion, then there are other implications for cosmogony. Many potential systems of terrestrial-type planets may be disrupted by a gas giant in their midst; we would owe our existence to the possibly fortuitous circumstance that only two cores in our own system, Jupiter and Saturn, grew large enough to accrete gas. It is not clear whether this fortunate outcome was purely stochastic, or influenced by the mass and density distribution of the solar nebula. There should also be a population of interstellar 'Jupiters' that were ejected from forming systems.

Our model can account, at least qualitatively, for the large eccentricities observed for 70 Vir B and HD114762 B. It cannot explain the other planets unless some additional process acted to circularize their orbits after they were scattered inwards. Processes that could reduce eccentricities include tidal interactions with the circumstellar disk¹⁷, drag of nebular gas enhanced by the planet's gravity¹⁸, and accretion of a portion of the nebula by the planet after migration to a new orbit. Any or all of these might explain the low eccentricity observed for 47 UMa B, whereas the disks about 70 Vir and HD114762 may have had insufficient mass or lifetimes too short to accomplish such damping.

It is more difficult to account for very close orbits such as that of 51 Peg B. In principle, gravitational encounters can produce an eccentric orbit with very small periastron, which could then be circularized by the processes mentioned above, or by tidal dissipation within the star and the planet (Mayor and Queloz¹ derived a timescale of a few billion years for such tides to damp the eccentricity of 51 Peg B). The subsequent discoveries of other planets in similar orbits make this explanation unlikely, as such 'star-grazing' orbits are rarely produced in our simulations. Observations suggest that a few per cent of solar-type stars may possess large planets at distances of 0.1 AU or less. Radial velocity surveys are most sensitive to close orbits, but it is unclear whether this bias can explain their apparent abundance. At present, there are appear to be two classes of orbits: those that are circular and very near their stars, and more distant ones that may have significant eccentricities. If this trend is confirmed by additional discoveries, it would suggest that two mechanisms for planetary migration were effective. The close orbits may be the product of tidal locking to an evolving disk, while the eccentric ones are due to gravitational scattering.

One test of the two mechanisms is the existence of other planetary companions. Tidal locking does not require that other planets exist around a given star, but if they do, their orbits should be coplanar with the inner one. Gravitational scattering implies that a star with one close planet should have at least one other planet of comparable mass in a distant eccentric orbit; the two orbits may have high mutual inclination. Detection of such a distant companion will be difficult. Radial velocity perturbations vary in amplitude as $a^{-1/2}$, but their period increases as $a^{3/2}$. In the example of Fig. 1, the effect of the outer planet would be about 1/4 as large as that of the inner one, but its period more than 50 times longer. Detection of a planet by radial velocity variations requires a length of observation comparable to the orbital period. As high-precision radial velocity surveys have been conducted for about a decade, it is not surprising that the only planets discovered to date are close to their stars. More distant companions should be found eventually by long-term monitoring of stellar radial velocities. They may also be revealed by their perturbations on the orbits of the inner planets, by astrometry, or by direct imaging.

Note added in proof: Cochran *et al.*¹⁹ have discovered another extra-solar planet with an eccentric orbit. This companion to the

star 16 Cygni B has $a = 1.7 \text{ AU}$, $e = 0.65$, and mass at least 1.6 times that of Jupiter. Rasio and Ford²⁰ have independently developed a model for planetary migration similar to ours. □

Received 8 April; accepted 30 October 1996.

- Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
- Latham, D. W., Mazeh, T., Stefanik, R. P., Mayor, M. & Burki, G. *Nature* **339**, 38–40 (1989).
- Marcy, G. W. & Butler, R. P. *Astrophys. J.* **464**, L147–L152 (1996).
- Butler, R. P. & Marcy, G. W. *Astrophys. J.* **464**, L153–L156 (1996).
- Lissauer, J. *Icarus* **69**, 249–265 (1987).
- Pollack, J. *et al.* *Icarus* (in the press).
- Lin, D. N. C., Bodenheimer, P. & Richardson, D. C. *Nature* **380**, 606–607 (1996).
- Boss, A. *Science* **267**, 360–362 (1995).
- Boss, A. *Nature* **379**, 397–398 (1995).
- Beckwith, S. & Sargent, A. *Nature* **383**, 139–144 (1996).
- Gladman, B. *Icarus* **106**, 247–263 (1993).
- Chambers, J., Wetherill, G. W. & Boss, A. *Icarus* **119**, 261–268 (1996).
- Farinella, P. *Moon Planets* **22**, 25–29 (1980).
- Everhart, E. in *Dynamics of Comets: Their Origin and Evolution* (eds Carusi, A. & Valsecchi, G.) 185–202 (Reidel, Dordrecht, 1985).
- Wetherill, G. W. *Nature* **373**, 470 (1995).
- Wetherill, G. W. *Astrophys. Space Sci.* **212**, 23–32 (1994).
- Ward, W. R. *Icarus* **73**, 330–348 (1988).
- Takeda, H., Matsuda, T., Sawada, K. & Hayashi, C. *Prog. Theor. Phys.* **74**, 272–287 (1985).
- Cochran, W. D., Hatzes, A. P., Butler, R. P. & Marcy, G. W. *Bull. Am. Astron. Soc.* **28**, 1111 (1996).
- Rasio, F. A. & Ford, E. B. *Science* (in the press).

ACKNOWLEDGEMENTS. We thank D. R. Davis and W. K. Hartmann for discussions and comments. This work was supported by NASA Planetary Geology and Geophysics Program and the Italian Space Agency.

CORRESPONDENCE should be addressed to S.J.W. (e-mail: sjw@psi.edu).

Quantum-size effects in the thermodynamic properties of metallic nanoparticles

Y. Volokitin*, J. Sinzig*, L. J. de Jongh*, G. Schmid†, M. N. Vargaftik* & I. I. Moiseev‡

* Kamerlingh Onnes Laboratory, Leiden University, PO Box 9506, NL 2300 RA Leiden, The Netherlands

† Institut für Anorganische Chemie, Universität GH Essen, Universitätsstrasse 5–7, D-45117 Essen, Germany

‡ Kurnakov Institute of General and Inorganic Chemistry, Academy of Science, 117071 Moscow V-71, Russia

THE properties of nanometre-scale metallic particles differ from those of the same material^{1,2} in bulk. Conduction electrons, because of their wave-like nature, can have only certain discrete values of kinetic energy or wavelength. Such ‘quantum-size’ effects have been observed in two-dimensional electron gases in semiconductors^{3,4}, and in atomic-scale metallic point contacts⁵. Also present are ‘Coulomb-charging’ effects: these are purely classical in origin, and occur when the energy required to add one electron to a conducting sphere exceeds the mean thermal energy $k_B T$. Thermal fluctuations in the total charge on the particle are then suppressed⁶. In theory, the combination of quantum-size and Coulomb-charging effects should cause the properties of small metallic particles to depend sensitively on whether they have an odd or even number of electrons⁷. Odd–even effects have been observed in experiments on tunnelling between discrete electronic levels of single metal particles⁸, but their influence on thermodynamic properties remains to be demonstrated. Here we report measurements of the heat capacity and electronic magnetic susceptibility of small metallic clusters. Our results show definitive evidence for odd–even effects, thus confirming that quantum and classical size effects strongly influence the thermodynamic properties of small particles.

Strong deviations from bulk metal behaviour are predicted for the electronic susceptibility and specific heat, χ_{el} and C_{el} , when

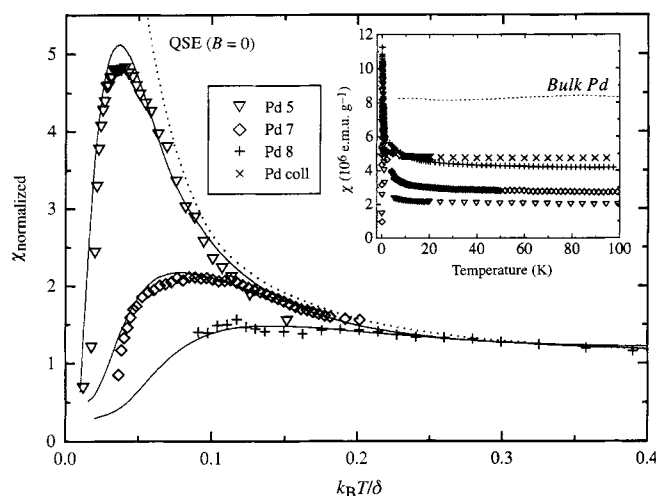


FIG. 1 Electronic magnetic susceptibility χ_{el} of the Pd clusters in the quantum-size regime. In the range 2–300 K the measurements were made with a commercial SQUID magnetometer; in the range 50 mK to 4 K, a SQUID a.c.-susceptometer made by the authors was used. The estimated diamagnetism of the ligand shells (a few per cent) has been subtracted from the susceptibility data. The inset shows the high- T behaviour, compared to that of bulk Pd. The low- T data have been normalized by dividing the susceptibilities by their high- T values, which will correspond to the electronic contributions unaffected by the presence of quantum gaps near to E_F . Temperature is scaled by the average energy gap value δ appearing in the theory. Dotted curve, predictions of the QSE model in zero field; solid curves, fits to the QSE model in field, yielding δ/k_B values of 12.5 K, 5 K and 2.0 K for Pd⁵, Pd⁷ and Pd⁸, respectively. To explain the observed susceptibility maxima, interactions between spins on neighbouring clusters have to be invoked, yielding internal fields of the order of 0.2 T, namely $\mu_B B / \delta = 0.025, 0.06, 0.15$ for Pd⁵, Pd⁷, and Pd⁸.

entering the quantum-size regime^{1,2}. For temperatures $T \lesssim 0.5 \delta/k_B$, where δ is the average electronic level spacing, C_{el} will drop to zero at a rate faster than linear, whereas χ_{el} will diverge or vanish for $T \rightarrow 0$ for the odd- or even-electron particles, respectively. A major experimental difficulty has been the unavailability of samples containing well characterized, monodisperse metal particles (that is, particles of the same size) in sufficiently high density. To date, size distributions of the order of 10% or higher could not be avoided, and such distributions will mask the effects expected. In addition, it has been pointed out⁹ that, even in an assembly of monodisperse particles, the precise electronic energy level structure will vary from particle to particle, owing to the extreme sensitivity to slight irregularities in the surface structure (on the atomic scale, a particle will always show surface roughness). When averaging over an assembly of metal particles, a distribution of energy levels has to be considered. This problem may be treated by random matrix theory, as developed for classifying nuclear energy levels and first applied to small particles by Gor'kov and Eliashberg⁹. Depending on the symmetry properties of the hamiltonian, only three possible distributions may occur, namely the orthogonal, unitary and symplectic distributions. The theory has been advanced by Denton *et al.*¹⁰, who have given quantitative predictions for the thermodynamic behaviour.

The materials studied here belong to a class of molecular metal clusters in which the metal cores are members of a series of ‘magic-number’ clusters obtained by surrounding an atom by successive shells of atoms of its kind^{11,12}. The (magic) number of atoms per particle obtained in this way increases as 13, 55, 147, 309, 561 and so on, corresponding to one-, two-, three-, four- and five-shell clusters. The four-shell cluster is found in the compound Pt₃₀₈Phen₅₆O₃₀ (ref. 13), whereas the five-, seven- and eight-shell clusters have been realized for Pd, namely in Pd₅₆₁Phen₃₆O₂₀₀, Pd₁₄₁₅Phen₅₄O₁₀₀₀ and Pd₂₀₅₇Phen₇₈O₁₆₀₀ (refs 14, 15). In addition, the related five-shell Pd cluster Pd₅₆₁Phen₄₀Ac₁₈₄O₃₇₀ has been

made¹⁶. Here Phen is phenanthroline and Phen* and Phen^{bu} are phenanthroline derivatives. Because these materials are stoichiometric chemical compounds, each sample contains the same kind of cluster molecules, so that the metal cores are ideally monodisperse. (We checked our samples by high-resolution electron microscopy). From previous work^{12,17–19} it has been established that the ligand shell profoundly affects the surface metal atoms of the clusters, but that the inner-core metal atoms already behave as small pieces of bulk metal¹⁸. An even larger core size is obtained in a Pd colloid stabilized by sodium sulphinate, with a narrow size distribution of about 5–10% around a diameter $d \approx 15$ nm ($\sim 1.25 \times 10^5$ atoms per cluster). The cluster compounds are denoted below by their number of shells, that is Pd⁵, Pd⁷, Pd⁸, and the colloid by Pd_{coll}. Samples weighed typically 20–80 mg.

The measured susceptibility data in the range 2–300 K (Fig. 1 inset) agree well with previously published results¹⁹, and demonstrate the evolution with increasing size to the behaviour of bulk Pd, for which χ_{el} is known to be strongly enhanced with respect to the free-electron Pauli susceptibility, owing to correlations between the electrons. The slow size-evolution towards the bulk has been explained¹⁹ in terms of surface effects, leading to a size-dependent reduction of both the density of states (DOS) at the Fermi energy E_F and the Stoner-enhancement factor. Of interest for the present work is the upturn in susceptibility often observed at the lowest temperatures¹⁹. Analysis in terms of a Curie contribution reveals that it corresponds to one spin for every two or three cluster molecules. As this number varies somewhat for different samples of the same compound, its origin has been attributed to magnetic impurities. Bulk Pd is reputedly sensitive to the presence of foreign magnetic ions²⁰, owing to its above-mentioned high magnetic polarizability. However, we note that the effects are most pronounced in Pd⁵, much less in Pd⁷ and Pd⁸, and essentially absent in the Pd colloid (see Fig. 1 and ref. 19). Why should the sensitivity to such impurities depend on size?

The answer appears to be provided by the combination of low-temperature ($T < 2$ K) specific heat and susceptibility data presented here, from which we conclude that the observed 'spurious' phenomena in the susceptibility should in fact be explained in terms of quantum-size effects (QSE) of the Pd metal cores, a fraction of which are carrying an odd number of electrons. The observed size dependence is then immediately understood: the upturn in χ_{el} , due to the Curie-type contribution from the odd-electron particles, will only develop for temperatures $T \lesssim 0.5 \delta/k_B$ (the contribution from the even-electron particles vanishes for $T \rightarrow 0$). Since δ will decrease with the inverse of the particle volume, the appearance of the QSE is shifted to lower temperatures with increasing size. To explain the maxima in the experimental χ_{el} we should take into account the effects of magnetic interactions between spins on neighbouring particles, which will be of dipolar origin with in addition a possible contribution due to super-exchange. The latter will involve two ligand shells and will therefore be quite weak, of the same order as the dipolar interaction which may be estimated from μ_B^2/r^3 (where r is the interparticle distance). With $r = 0.5$ nm, a value of 10^{-6} eV is obtained. When half of the clusters carry a spin, each spin will interact with about six nearest neighbours, so that a given spin will 'feel' an internal field of ~ 0.1 tesla. In our materials the cluster molecules are randomly packed, as in a glass. There will thus be randomness in the magnetic intercluster interactions, resulting in short-range magnetic order at the lowest temperatures, as in a magnetic glass. The internal fields provide an explanation for the pronounced downturn in the χ_{el} data shown in Fig. 1. In the QSE

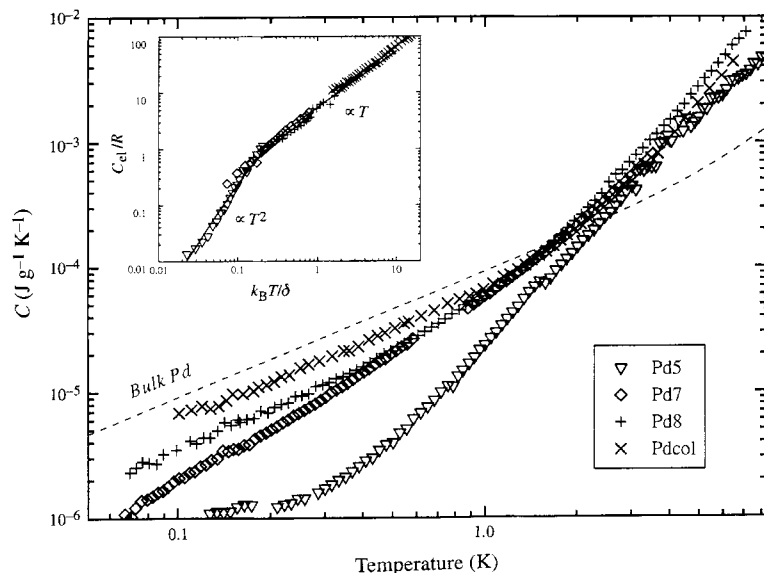


FIG. 2 Measured specific heat C of the same Pd clusters as in Fig. 1 compared to bulk Pd. Data below 1 K were obtained in the same ^3He – ^4He refrigerator used for the susceptibility measurements. Above 3 K, the T^3 -dependence expected on basis of the lattice vibrations is followed (phonon contributions from clusters and ligands). Below 1 K the observed specific heat may be attributed to the electronic contributions from the Pd cluster cores. Inset, the electronic part (data points) obtained after subtracting the phonon contributions. The fits to the curve predicted by the QSE theory (solid line) yield δ/k_B values of 12 K, 4.5 K, 3.0 K and 0.06 K for Pd⁵, Pd⁷, Pd⁸ and Pd_{coll}, respectively. The transition from metallic behaviour ($C_{el} \propto T$) to quantum-size behaviour ($C_{el} \propto T^2$) is clearly seen.

model, the application of a magnetic field B of sufficient strength ($B \gtrsim 0.02 \delta/\mu_B$) leads to saturation of the magnetization at low enough temperatures ($T \lesssim 0.02 \delta/k_B$) and thus to a vanishing differential susceptibility (the field lifts the spin-degeneracy of the odd-electron levels). The solid curves in Fig. 1 give the QSE predictions (discussed below) for the field-values mentioned in the legend. These values (0.1–0.3 T) are indeed of the order of the above-estimated internal fields. The influence of the internal fields should be most pronounced for the smallest particles, in agreement with experiment.

We now turn to the specific heat measurements, presented in Fig. 2 together with data on bulk Pd. The bulk specific heat can be decomposed into a linear term, coming from the conduction electrons ($C_{el} \propto T$), and a cubic term due to the lattice vibrations ($C_{vib} \propto T^3$). For our ligated clusters, a strong enhancement of the cubic term is seen, which can be attributed to the presence of the ligand shells. After subtracting the cubic terms, the remaining specific heat below 1 K for the clusters should be attributed to the electronic contribution from the metal cores, which is strongly size-dependent. C_{el} is almost absent for Pd⁵ and shows a clear linear term only for Pd_{coll}. However, the linear term is still ~ 20 – 30% lower than for the bulk, indicating a substantially lower average DOS at E_F even for Pd_{coll}, in agreement with the previous analysis¹⁹ of the susceptibility above 1 K. The lower DOS is attributed in part to surface effects, such as the lower coordination number and the metal–ligand interactions.

Studies of the magnetic-field dependence of the specific heat confirm the conclusions from the susceptibility data regarding the presence of unpaired spins on odd-electron particles. A strong field dependence is observed only for Pd⁵ (Fig. 3). Analysis of the broad anomalies, which develop in field owing to the Zeeman splitting of the energy levels, indicates a similar fraction of spins as do the susceptibility data. The Pd⁷, Pd⁸ and Pd_{coll} samples show much weaker field dependences of the specific heat (not shown here); the field effect becomes appreciable only for $T < 0.2$ K (even for a field of 6 T). The specific-heat data also show the

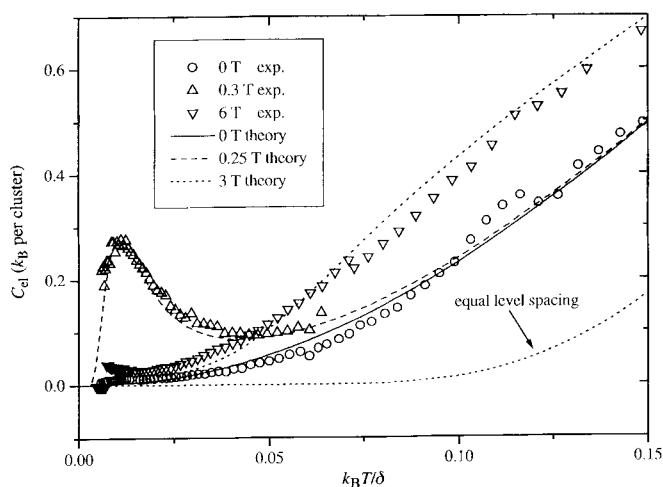


FIG. 3 Magnetic-field dependence of the electronic specific heat of Pd^5 . For low applied fields, $B < 0.1 \delta/\mu_B$, the data are in qualitative agreement with the QSE model described in the text. Owing to the limitations of the model, no quantitative agreement can be expected for fields in the tesla range. The theoretical curve for 3 T is therefore only shown as an indication of the behaviour in high fields. The dotted curve shows the prediction¹⁰ for the equal-level spacing in zero field.

presence of internal fields due to interparticle magnetic interactions, in particular for Pd^5 (below 0.2 K).

We are convinced that quantum-size effects in an ensemble of even-odd electron particles provide the only possible explanation for our data. The theoretical curves shown in Figs 1–3 have been obtained using an extension²¹ of the QSE model calculations¹⁰. Since in our experiments on Pd we are in the limit of weak field and weak spin-orbit coupling, we adopt the orthogonal distribution for the level separations between E_F and the two nearest higher levels. All other levels are taken as equidistant in order to minimize the time needed for the numerical computations. This approximation should be quite reasonable¹⁰, as at low temperature only these few levels are populated, whereas at high temperatures ($T \gg \delta/k_B$) the details of the level distribution become irrelevant. The fits were made assuming an equal number of odd- and even-electron particles. For both χ_{el} and C_{el} they agree to an average level separation δ/k_B of 12 ± 1 K, 5 ± 1 K and 2.5 ± 1.5 K for Pd^5 , Pd^7 and Pd^8 , respectively. We stress that both the susceptibility and specific-heat data may be fitted within the errors by the same values for δ , which is the only fitting parameter in the QSE model. (That the values of δ obtained do not scale fully with the volumes of the particles is attributed to the above-mentioned surface effects on the DOS at E_F .) For Pd_{coll} the δ value obtained (0.06 ± 0.02 K) is already so small that it could only be derived from the specific heat. It is of interest that the δ values are rather close to those obtained from the simple formula $4E_F/3N$, where N is the number of valence electrons in the cluster. With 10 electrons per Pd atom, and $E_F = 6.8$ eV, we obtain δ/k_B values of 19 K, 7.5 K, 5 K and 0.08 K for Pd^5 , Pd^7 , Pd^8 and Pd_{coll} .

The trends in the experimental data appear to match well the predictions of the model. In particular the T^2 -dependence of C_{el} predicted for the orthogonal ensemble in the QSE regime, agrees well with the data on Pd^5 . The latter clearly rule out the equal-level-spacing model, which predicts exponential decay of C_{el} for $T \rightarrow 0$ (Fig. 3). The observed T^2 -dependence is to our knowledge the first experimental evidence for the presence of level-repulsion in metal particles. Also, the field-dependent phenomena seen for Pd^5 appear to be qualitatively reproduced, given the limitations of the model, in which the orthogonal distribution is restricted to the two levels closest above E_F . Accordingly, in an applied field the model will start to fail as soon as the Zeeman splitting becomes comparable to δ (that is, for fields exceeding a few tesla). $\square$

Received 19 July; accepted 5 November 1996.

1. Perenboom, J. A. A., Wyder, P. & Meier, F. *Phys. Rep.* **78**, 173–292 (1981).
2. Halperin, W. P. *Rev. Mod. Phys.* **58**, 533–606 (1986).
3. van Wees, B. J. et al. *Phys. Rev. Lett.* **60**, 848–850 (1988).
4. Wharam, D. A. et al. *J. Phys. C* **21**, L209–L214 (1988).
5. Krans, J. V., van Ruitenbeek, J. M., Fisman, V. V., Yanson, I. K. & de Jongh, L. J. *Nature* **375**, 767–769 (1995).
6. Gorter, C. J., *Physica* **17**, 777–780 (1951).
7. Kubo, R. *J. Phys. Soc. J.* **17**, 975–986 (1962).
8. Ralph, D. C., Black, C. T. & Tinkham, M. *Phys. Rev. Lett.* **74**, 3241–3244 (1995).
9. Gor'kov, L. P. & Eliashberg, G. M. *Sov. Phys. JETP* **21**, 940–947 (1965).
10. Denton, R., Mühlischlegel, B. & Scalapino, D. J. *Phys. Rev. B* **7**, 3589–3607 (1973).
11. Schmid, G. (ed.) *Clusters and Colloids. From Theory to Applications* (VCH, Weinheim, 1994).
12. de Jongh, L. J. (ed.) *Physics and Chemistry of Metal Cluster Compounds* (Kluwer, Dordrecht, 1994).
13. Schmid, G., Morun, B. & Malm, J.-O. *Angew. Chem. Int. Edn Engl.* **28**, 778–780 (1989).
14. Schmid, G. *Polyhedron* **7**, 2321–2329 (1988).
15. Schmid, G. et al. *J. Am. Chem. Soc.* **115**, 2046–2048 (1993).
16. Vargaftik, M. N. et al. *J. Chem. Soc. Chem. Commun.* 937–939 (1985).
17. van Leeuwen, D. A. *Phys. Rev. Lett.* **73**, 1432–1435 (1994).
18. Mulder, F. M., Stegink, T. A., Thiel, R. C., de Jongh, L. J. & Schmid, G. *Nature* **367**, 716–718 (1994).
19. van Leeuwen, D. A., van Ruitenbeek, J. M., Schmid, G. & de Jongh, L. J. *Phys. Lett. A* **170**, 325–333 (1992).
20. Foner, S., Doclo, R. & McNiff, E. J. Jr *J. Appl. Phys.* **39**, 551–552 (1968).
21. de Jongh, L. J. & Sinzig, J. in *Localized and Itinerant Molecular Magnetism: From Molecular Assemblies to the Devices* (eds Coronado, E., Delhaes, P., Gatteschi, D. & Miller, J. S.) (Kluwer, Dordrecht, in the press).

ACKNOWLEDGEMENTS. We thank H. B. Brom and J. M. van Ruitenbeek for their interest. This work was part of the research program of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) which is supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). The support of the European Community under the HCM programme is acknowledged.

CORRESPONDENCE and requests for materials should be addressed to L.J.d.J.

Vegetation and soil feedbacks on the response of the African monsoon to orbital forcing in the early to middle Holocene

J. Kutzbach*, G. Bonan†, J. Foley* & S. P. Harrison‡

* Center for Climatic Research, University of Wisconsin-Madison, 1225 West Dayton Street, Madison, Wisconsin 53706, USA

† National Center for Atmospheric Research, PO Box 3000, Boulder, Colorado 80307-3000, USA

‡ Dynamic Palaeoclimatology, Lund University, Box 117, Lund, S-221 00, Sweden

Fossil pollen, ancient lake sediments and archaeological evidence from Africa indicate that the Sahel and Sahara regions were considerably wetter than today during the early to middle Holocene period, about 12,000 to 5,000 years ago^{1–4}. Vegetation associated with the modern Sahara/Sahel boundary was about 5° farther north, and there were more and larger lakes between 15 and 30° N. Simulations with climate models have shown that these wetter conditions were probably caused by changes in Earth's orbital parameters that increased the amplitude of the seasonal cycle of solar radiation in the Northern Hemisphere, enhanced the land–ocean temperature contrast, and thereby strengthened the African summer monsoon^{5–7}. However, these simulations underestimated the consequent monsoon enhancement as inferred from palaeorecords⁴. Here we use a climate model to show that changes in vegetation and soil may have increased the climate response to orbital forcing. We find that replacing today's orbital forcing with that of the mid-Holocene increases summer precipitation by 12% between 15 and 22° N. Replacing desert with grassland, and desert soil with more loamy soil, further enhances the summer precipitation (by 6 and 10% respectively), giving a total precipitation increase of 28%. When the simulated climate changes are applied to a biome model, vegetation becomes established north of the current Sahara/

Sahel boundary, thereby shrinking the area of the Sahara by 11% owing to orbital forcing alone, and by 20% owing to the combined influence of orbital forcing and the prescribed vegetation and soil changes. The inclusion of the vegetation and soil feedbacks thus brings the model simulations and palaeovegetation observations into closer agreement.

Studies with climate models suggest that major changes in vegetation can affect climate significantly: replacement of tundra by boreal forest causes warming^{8,9}, desertification^{10–12} and tropical deforestation^{13,14} cause drying. But could vegetation/soil feedbacks have enhanced the climatic response of monsoons to orbital forcing 9,000 to 6,000 years ago? A previous climate model experiment found that the response of precipitation to orbital forcing was increased by lowering the surface albedo of the Sahara from ~0.30 to ~0.25 to represent desert being replaced by grassland at 9,000 years ago². Whereas this study used a relatively low-resolution atmospheric model (5° latitude by 7.5° longitude, and 11 vertical levels) with a relatively simple land-surface model that characterized vegetation differences only in terms of differences of surface albedo, we use an atmospheric model of higher resolution (2.8° by 2.8° and 18 levels) and employ a land-surface model (LSM) with an explicit treatment of vegetation and soil processes^{15–17}. The atmospheric model is a modified version of the NCAR Community Climate Model version 2 (CCM2)^{18–20}. The LSM solves for the vegetation and ground temperatures that balance the surface energy budget, taking into account ecological differences among vegetation types (leaf and stem area, leaf optical properties, leaf dimensions, surface roughness, stomatal physiology, root profile) and hydraulic and thermal differences among soil types (porosity, hydraulic conductivity, matric potential, slope of retention curve, thermal conductivity, heat capacity, albedo).

We did four climate simulations: control (modern climate), R (changes in radiation at 6,000 years ago), RV (changes in radiation and vegetation) and RVS (changes in radiation, vegetation and soil). The atmospheric CO₂ concentration and the seasonal cycle of ocean surface temperature are prescribed at modern values in all four simulations. In R we change the Earth's orbital parameters from modern to 6,000 years ago: eccentricity from 0.167 to 0.187, axial tilt from 23.4° to 24.1°, and date of perihelion from early January to mid-September²¹. The change in the date of perihelion increases the amplitude of the seasonal cycle of Northern Hemisphere insolation by ~5%, leading to enhanced heating in northern summer–autumn and enhanced cooling in northern winter–spring. In RV we modify the vegetation in northern Africa between 15° N and 30° N from desert (100% bare ground) or semi-desert (10% shrub, 90% bare ground) (control and R) to grassland (80% grass, 20% bare ground). The northern limit of the prescribed grassland goes beyond the limit indicated by palaeobotanical evidence^{1–4} and was chosen to test the model's sensitivity to a large change². It would have been more accurate to prescribe a shrub/grass mixture rather than grassland, however the LSM does not include this category. The change to grassland increases the leaf and stem areas within the vegetated fraction, decreases the surface albedo slightly, modifies the sensible heat flux, and modifies the partitioning

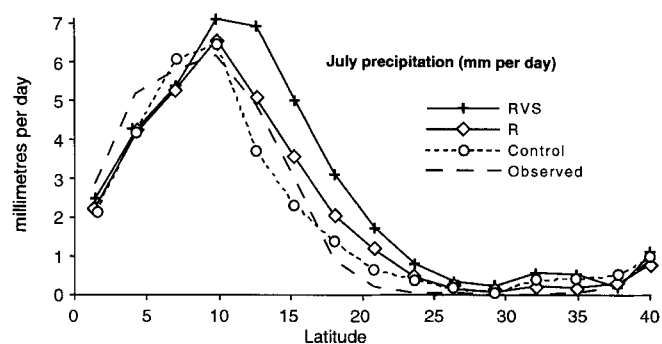


FIG. 1 July precipitation (mm per day) as a function of latitude and averaged longitudinally for the sector 0–50° E, for the control, experiment R and experiment RVS, and modern observations^{30,31}. Longitudinal averages are appropriate because the climate and vegetation of northern Africa are rather uniform longitudinally.

of the total latent heat flux into evaporation from soil, evaporation from vegetation, and transpiration. In RVS, we prescribe both the increased grassland and a more loamy soil to reflect the expected long-term increase in organic content of the soil under grassland. There is evidence of such a Holocene soil development from several locations in the Sahara^{22,23}. This change increases the available water capacity of the soil (from 0.22 m water per metre of soil in the control, R and RV to 0.24 m, an increase of ~10%) and decreases the rate at which the soil drains. Consistent with the higher organic content, we also change soil colour so that albedo decreases from 0.25 to 0.22 (Table 1).

Each of the simulations was run for 5.5 years. Systematic trends in soil moisture were small and generally similar in the four simulations after the first 1.5 years of spin-up and there were no discernible trends in other variables (such as temperature, precipitation). Although longer spin-up periods have been proposed²⁴, we believe that 1.5 years is adequate for these simulations. The control simulates July precipitation in good agreement with observations (Fig. 1).

Radiative forcing (R minus control) increases July insolation by ~20 W m⁻² (Fig. 2), leading to increases in net radiation at the

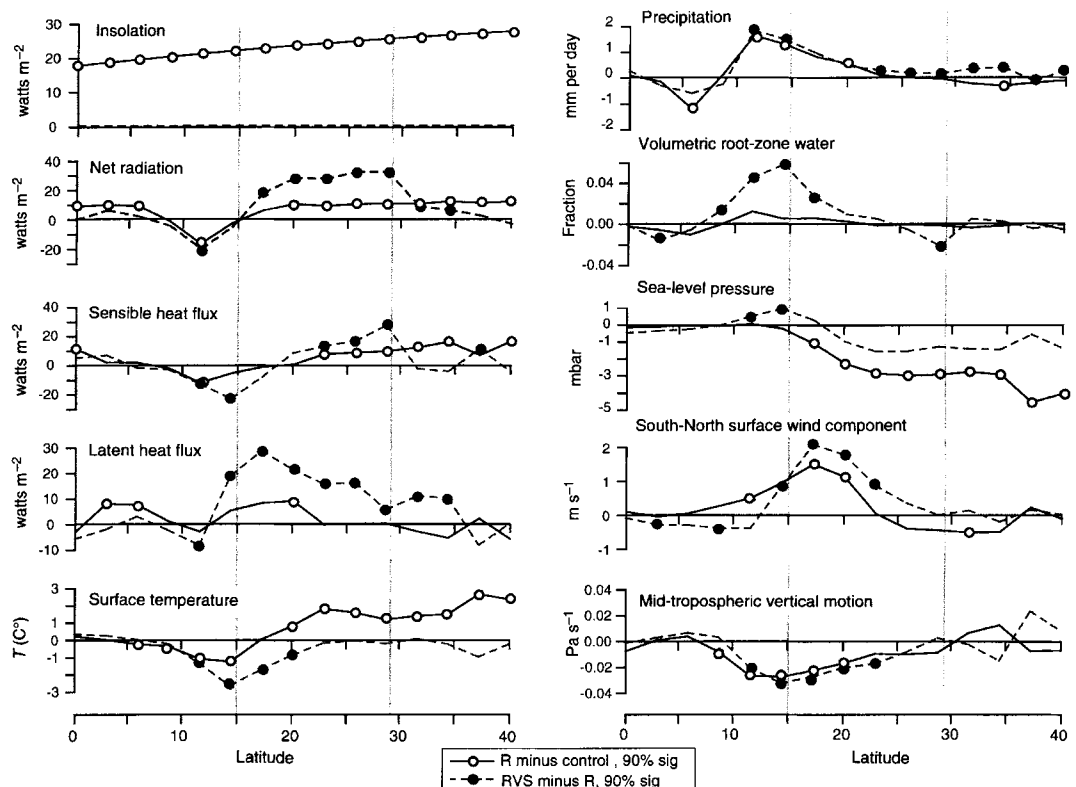
TABLE 1 Annual average climate, vegetation and soil variables for four simulations (15° N–22° N, 0–50° E)

Variable	Experiment			
	Control	R	RV	RVS
Insolation (top of atmosphere) (W m ⁻²)	399	398	398	398
Surface albedo (fraction)	0.25	0.25	0.25	0.22
Net surface radiation (W m ⁻²)	94	95	98	104
Sensible heat flux (W m ⁻²)	56	53	54	57
Evapotranspiration (W m ⁻²)	38	42	44	47
(Soil/canopy evaporation, W m ⁻²)	(35)	(39)	(35)	(39)
(Canopy transpiration, W m ⁻²)	(3)	(3)	(9)	(8)
Precipitation (mm per day)	1.28	1.43	1.51	1.63
Precipitation minus evapotranspiration (mm per day)	-0.03	-0.01	-0.02	0.00
Volumetric root-zone water (fraction)	0.18	0.18	0.18	0.20
Near-surface temperature* (°C)	28.7	27.5	27.3	27.5
Near-surface specific humidity* (g kg ⁻¹)	6.2	6.8	7.1	7.4
Total cloud (fraction)	0.33	0.37	0.36	0.37
Precipitation (15° N–22° N; fraction of control)	1.00	1.12	1.18	1.28
Precipitation (23° N–30° N; fraction of control)	1.00	0.99	0.96	1.13

R, radiation 6,000 years ago; RV, radiation and changed vegetation 6,000 years ago; RVS, radiation and changed vegetation and soils 6,000 years ago, averaged over the southern half of the region of modified vegetation/soil in RV and RVS (15° N–22° N, 0–50° E). The annual average is based on the final four years of simulation. At the bottom of the table, changes in precipitation (fraction) are shown, averaged over the southern and northern halves of the region of modified vegetation/soil. Values in bold denote that the differences R – control, RV – control, RVS – control are statistically significant at the 95% level.

* Near-surface refers to the lowest level of the model's atmosphere, 0.992σ; this corresponds roughly to a height of 50 m.

FIG. 2 Changes in simulated climatic conditions for July between experiment R and the control (solid lines), representing the response to orbital forcing, and between experiment RVS and experiment R (dashed lines), representing the response to vegetation/soil forcing; the changes are depicted as a function of latitude and are longitudinal averages, 0–50° E, averaged for the last four years of simulation. The symbols on each graph indicate that the change is significant at the 90% level (based on the *t*-test statistic of the difference in mean values compared to the model's own year-to-year variability). Vertical lines indicate the region of modified vegetation/soil in experiment RVS. In the chart illustrating surface winds, positive values indicate southerly wind components. In the chart illustrating vertical motion, more upward motion is negative and 0.01 Pa s^{-1} corresponds to $\sim 1.5 \text{ mm s}^{-1}$.



surface ($\sim 10 \text{ W m}^{-2}$), in sensible heat flux ($\sim 10 \text{ W m}^{-2}$) and in surface temperature (~ 1 to 2°C north of 20°N). The enhanced monsoon circulation is indicated by decreased sea-level pressure (~ -2 to -3 mbar), increased southerly wind ($\sim 1 \text{ m s}^{-1}$) and increased mid-troposphere upward air motion (Fig. 2). Precipitation increases by 1–2 mm per day between 10°N and 20°N (Fig. 2), representing a 2° latitude northward extension of the summer monsoon (Fig. 1). In the region of enhanced rainfall, soil water increases slightly, evapotranspiration increases ($\sim 10 \text{ W m}^{-2}$) and cloudiness increases, leading to a decrease in temperature (about -1°C). The response of precipitation to orbital forcing in this version of CCM2 is substantially weaker than the response obtained previously with CCM1 (ref. 25), and also substantially less than is required to account for the observations of mid-Holocene moist environments. There are differences in both resolution and subgrid-scale parameterizations between these two models, so it is difficult to pinpoint the cause of their different climatologies and different monsoonal responses to orbital forcing^{26,27}.

Vegetation/soil forcing (RVS minus R) enhances the summer monsoon significantly compared to orbital forcing alone (Fig. 2). The climatic response to vegetation/soil forcing equals or exceeds the response to orbital forcing in magnitude, and both kinds of forcing produce similar changes with latitude. The surface net radiation increases by 25 W m^{-2} , latent heat flux (evapotranspiration) by $\sim 20 \text{ W m}^{-2}$, and soil water by more than 10%. There is additional lowering of sea-level pressure, increased southerly wind, and increased upward air motion (Fig. 2). The change in July precipitation, which is greatest over the southern half of the modified region, equals that due to orbital forcing alone (Fig. 2) and represents an additional 2° latitude northward extension of monsoon rains (Fig. 1). The change in July precipitation is statistically significant at most latitudes from 13°N to 40°N , and changes in other variables are statistically significant over more limited domains (Fig. 2). Whereas the orbitally induced enhancement of precipitation (R-control) occurs only in the period July–September, coincident with the maximum in orbital forcing, the enhance-

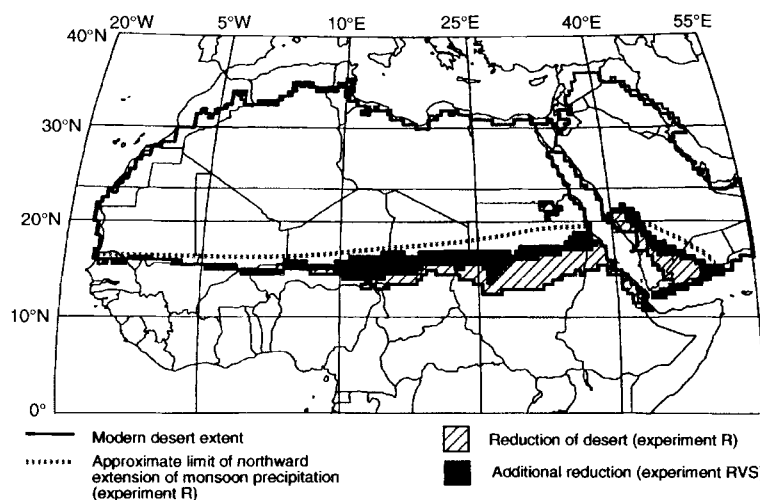
ment of precipitation due to vegetation/soil forcing (RVS minus R) occurs throughout the summer monsoon, May–September.

The hydrological response to vegetation/soil forcing is a strong function of the model's climate for the control and R. In the northern half of the modified region, 23°N – 30°N , precipitation is very low in the control and in R (Figs 1 and 2), and therefore the response of soil moisture and precipitation in RVS is small (Fig. 2). In the southern half of the modified region, 15°N – 22°N , where precipitation in the control exceeds 1 mm per day and where the response of precipitation in R is largest (Figs 1, 2), the response of the hydrological cycle to vegetation/soil forcing is considerable (Figs 1, 2).

Both the increased available water capacity of the soil and the decreased surface albedo in RVS, compared to RV, appear to be important. Root-zone soil water increases 10% in RVS, relative to RV, R and the control, and the model's evapotranspiration is sensitive to this increase (Table 1). The lowered surface albedo in RVS accounts for much of the increased net radiation between RVS and R (9 W m^{-2}); the corresponding increase between RV and R is only 3 W m^{-2} . This increase in available water and in net radiation in RVS, compared to R, helps drive the increased evapotranspiration (5 W m^{-2}) contributed by increased canopy transpiration; the corresponding increase between RV and R is only 2 W m^{-2} . In response to the increased evapotranspiration, near-surface specific humidity increases from 6.8 g kg^{-1} in R to 7.1 g kg^{-1} in RV to 7.4 g kg^{-1} in RVS. The increased humidity, low-level flow convergence, and upward vertical motion (Fig. 2) combine to enhance precipitation: R, 12% increase, RV, 18% increase, RVS, 28% increase, all compared to the control (Table 1).

Could the prescribed vegetation/soil changes have increased precipitation sufficiently in the southern half of the region to be self-maintaining? We begin to address this question by forcing a physiologically based biome model²⁸ with the output from the climate model. The four-year simulations are too short to provide an accurate climatology for the biome estimates, but nevertheless illustrate the direction of change. Compared to the modern

FIG. 3 Location of desert as simulated by a physiologically based biome model²⁸ driven by output from experiments R and RVS (mean monthly temperature, precipitation, solar radiation), compared to modern. The hatched line marks the approximate limit of northward extension of monsoon precipitation in experiment R. Slant hatching indicates the reduction in the area of desert in experiment R; dark shading indicates the additional reduction in experiment RVS. The shrinking desert is replaced by xerophytic woods/scrub and warm grass/shrub vegetation.



control, R produces a larger area of xerophytic woods/scrub and warm grass/shrub vegetation so that the desert area in the region 0–50° E, 0–30° N, is reduced by 11%; desert is reduced by 16% in RV and 20% in RVS (Fig. 3). The vegetation simulated by RVS is in better agreement with palaeovegetation than that simulated by R (refs 1–4), especially around Lake Chad and eastward into the Sudan where both the simulation and the palaeoenvironmental observations^{1–4} indicate vegetation 3° to 5° north of modern limits. None of the simulations produce vegetation in northwest Mali and northeast Sudan as indicated by observations^{1–4}. However, it is possible that the northernmost sites record local palaeovegetation rather than a broad continent-wide extension.

Our results and earlier studies^{2,7–17} suggest that climate models need to include biospheric processes involving both vegetation and soils. Other factors that might further enhance the climatic response to orbital forcing include the extensive lakes and marshes that existed in northern Africa in the middle Holocene (refs 1–4, and M. T. Coc, personal communication, and changes in ocean circulation that could influence sea surface temperatures. Our results also have implications for the accurate simulation of possible future climates. Climate models simulate changes in African monsoon precipitation in response to increased greenhouse gases²⁹, but these models will need to include possible vegetation and soil feedbacks to assess more accurately the magnitude of future climate change. □

28. Prentice, I. C. et al. *J. Biogeogr.* **19**, 117–134 (1992).

29. Mitchell, J. F. B., Manabe, S., Meleshko, V. & Tokioka, T. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T., Jenkins, G. J. & Ephraums, J. J.) 131–172 (Cambridge Univ. Press, Cambridge, 1990).

30. Legates, D. R. & Willmott, C. J. *Theor. Appl. Climatol.* **41**, 11–21 (1990).

31. Legates, D. R. & Willmott, C. J. *Int. J. Climatol.* **10**, 111–127 (1990).

ACKNOWLEDGEMENTS. We thank P. Behling, R. Selin and J. Kruepke for assistance with model simulations and graphics, and M. Kennedy for preparing the manuscript. This research was supported by grants to the University of Wisconsin–Madison by the NSF Climate Dynamics Program, and the Department of Energy, by the National Center for Atmospheric Research, which is sponsored by the NSF, and by awards to Lund University from the European Community, the Swedish Natural Science Research Council, and the Crafoord Fund.

CORRESPONDENCE should be addressed to J. K. (e-mail: jek@facstaff.wisc.edu).

Leading-edge vortices in insect flight

Charles P. Ellington, Coen van den Berg*, Alexander P. Willmott* & Adrian L. R. Thomas*

Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

INSECTS cannot fly, according to the conventional laws of aerodynamics: during flapping flight, their wings produce more lift than during steady motion at the same velocities and angles of attack^{1–5}. Measured instantaneous lift forces also show qualitative and quantitative disagreement with the forces predicted by conventional aerodynamic theories^{6–9}. The importance of high-lift aerodynamic mechanisms is now widely recognized but, except for the specialized fling mechanism used by some insect species^{1,10–13}, the source of extra lift remains unknown. We have now visualized the airflow around the wings of the hawkmoth *Manduca sexta* and a 'hovering' large mechanical model—the flapper. An intense leading-edge vortex was found on the downstroke, of sufficient strength to explain the high-lift forces. The vortex is created by dynamic stall, and not by the rotational lift mechanisms that have been postulated for insect flight^{14–16}. The vortex spirals out towards the wingtip with a spanwise velocity comparable to the flapping velocity. The three-dimensional flow is similar to the conical leading-edge vortex found on delta wings, with the spanwise flow stabilizing the vortex.

Several studies have visualized the airflow around flapping

Received 12 June; accepted 24 October 1996.

- Street, F. A. & Grove, A. T. *Nature* **261**, 385–390 (1976).
- Street-Perrott, F. A., Mitchell, J. F. B., Marchand, D. S. & Brunner, J. S. *Trans. R. Soc. Edinb.* **81**, 407–427 (1990).
- Petit-Maire, N. & Riser, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 45–61 (1981).
- Jolly, D., Harrison, S. P., Dammati, B. & Bonnefille, R. *Quat. Sci. Rev.* (in the press).
- Kutzbach, J. E. & Otto-Bliesner, B. L. *J. Atmos. Sci.* **39**, 1177–1188 (1982).
- Kutzbach, J. E. & Guetter, P. J. *J. Atmos. Sci.* **43**, 1726–1759 (1986).
- Mitchell, J. F. B., Grahame, N. S. & Needham, K. H. *J. Geophys. Res.* **93**, 8283–8303 (1988).
- Bonan, G. B., Pollard, D. & Thompson, S. L. *Nature* **359**, 716–718 (1992).
- Foley, J., Kutzbach, J. E., Coe, M. T. & Lewis, S. *Nature* **371**, 52–54 (1994).
- Charnay, J. G., Quirk, W. J., Chow, S. H. & Kornfeld, J. J. *J. Atmos. Sci.* **34**, 1366–1385 (1977).
- Laval, K. & Picon, L. *J. Atmos. Sci.* **43**, 2418–2429 (1986).
- Xue, Y. & Shukla, J. *J. Clim.* **6**, 2232–2245 (1993).
- Henderson-Sellers, A. et al. *J. Geophys. Res.* **98**, 7289–7315 (1993).
- Nobre, C. A., Sellers, P. J. & Shukla, J. *J. Clim.* **4**, 957–988 (1991).
- Bonan, G. B. *J. Geophys. Res.* **99**, 25803–25818 (1994).
- Bonan, G. B. *J. Geophys. Res.* **100**, 2817–2831 (1995).
- Bonan, G. B. *J. Clim.* **8**, 2691–2704 (1995).
- Hack, J. J. et al. *NCAR Technical Note TN-382+STR*, 1–108 (National Center for Atmospheric Research, Boulder, CO, 1993).
- Hack, J. J., Boville, B. A., Kiehl, J. T., Rasch, P. J. & Williamson, D. L. *J. Geophys. Res.* **99**, 20785–20813 (1994).
- Kiehl, J. T. *J. Geophys. Res.* **99**, 23107–23115 (1994).
- Berger, A. *J. Atmos. Sci.* **35**, 2362–2367 (1978).
- Haynes, C. V. *Natl. Geogr. Soc. Res. Rep.* **19**, 269–341 (1985).
- Petit-Maire, N. & Riser, J. *Sahara ou Sahel? Quaternaire récent du Bassin de Taoudeni (Mali)* (Lamy, Marseille, 1983).
- Milly, P. C. D. & Dunne, K. A. *J. Clim.* **7**, 506–526 (1994).
- Kutzbach, J. E. et al. *Quat. Sci. Rev.* (in the press).
- Sperber, K. R., Hameed, S., Potter, G. L. & Boyle, J. S. *PCMDI Rep. No. 12*, UCRL-ID-11532 (Lawrence Livermore National Laboratory, Livermore, CA, 1993).
- Boyle, J. S. *J. Clim.* **6**, 796–815 (1993).

* Present addresses: Faculty of Human Movement Sciences, Vrije Universiteit, Van der Boerhorststraat 9, 1081 BT Amsterdam, The Netherlands (C.v.d.B.); Kawachi Millibioflight Project, Japan Science and Technology Corporation (JST), Park Building 3F, 4-7-6 Komaba, Meguro-ku, Tokyo 153, Japan (A.P.W.); Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK (A.L.R.T.).

insect wings in an attempt to identify high-lift mechanisms^{17–22}. These studies have used planar or sheet illumination with smoke or particles to visualize cross-sections of the flow. The three-dimensional flow field can be reconstructed from a series of such images, but flow normal to the planes is difficult to detect. We therefore used stereophotography to record the three-dimensional flow of smoke around the wings of the hawkmoth *Manduca sexta* tethered in a windtunnel; to our knowledge, this is the first study in which medium and high speeds were investigated in addition to low speeds. *Manduca* has a wing motion typical of many insects, and the downstroke is primarily responsible for weight support. A vertical smoke rake was positioned on the wind-tunnel axis, and different spanwise positions were examined by moving the tethered moth relative to the smoke (described in detail elsewhere²³).

Figure 1a shows the flow just after the middle of the downstroke, with the plane of smoke filaments halfway along the left forewing (airspeed is 3.7 m s^{-1}). The flow separates at the leading edge and reattaches to the upper surface in the posterior half of the wing, enclosing a leading-edge vortex. Results with the smoke at other spanwise positions show a gradual enlargement of the vortex towards the wingtip. It is reasonable well structured until the outer quarter of the wing, where it breaks away and rolls up into a large diameter tip vortex. As smoke reattaches behind the leading-edge vortex, it changes direction abruptly and flows towards the wingtip and into the tip vortex. The tip vortex runs back to the starting vortex shed at the beginning of the downstroke, which is not visualized with the smoke rake at this location. The tip vortex is most clearly seen early in the downstroke, where the wingtip cuts through the smoke. Figure 1b is a stereo drawing that traces the path of the leading-edge/tip vortex.

The leading-edge vortex is a region of low pressure above the wing and thus will augment the lift force. The extra lift can also be explained as an enhanced circulation around the wing due to the vortex; the circulation is a measure of the velocity difference above and below the wing. The two explanations—pressure and circulation—are equivalent. Most of the unsteady aerodynamic mechanisms that have been proposed for insect flight feature lift enhancement by a leading-edge vortex^{10–16,24,25}, and experiments on rigid, model wings have confirmed that they are potentially effective for two-dimensional wing motions: that is, linear translation with rotation about a spanwise axis (reviewed in ref. 16). Given the size and speed of the wings (a Reynolds number of $\sim 10^3$) and their thin leading edge, flow separation at the leading edge and subsequent reattachment of a large, laminar vortex is almost inevitable.

However, evidence for the vortex from earlier insect studies is limited; the vortex is either not present or otherwise is much smaller than expected^{17–22}. It has even been suggested that leading-edge separation is prevented by wing surface microstructure or deformation²¹. We observed the leading-edge vortex at all flight speeds, from 0.4 to 5.7 m s^{-1} (Fig. 1c). It is very small, only a fraction of the chord, at 0.4 m s^{-1} ; note that the wake is unstable and quickly breaks down at that speed. Studies that failed to observe the vortex used particle visualization in still

air^{20,21}; the experimental technique was probably inadequate to see the small vortices. None of the earlier studies investigated the flow at medium and high speeds. The size of the vortex increases markedly with speed for *Manduca*, until at 5.7 m s^{-1} it extends over the entire chord. At all speeds, the size increases significantly during the course of the downstroke.

Although these results clearly reveal a leading-edge vortex, they do not identify the aerodynamic mechanism responsible for its creation: the vortex is a common feature of most high-lift mechanisms that have been postulated for insect flight. These mechanisms can be divided conveniently into those applicable to the translational phase of the wingbeat, when the fundamental flapping motion imparts a curvilinear motion to the wing elements, and those relevant to the rotational phase at either end of the wingbeat, when the angle of attack changes quickly in preparation for the next half-stroke¹⁶. Dynamic stall, or delayed stall, is the mechanism appropriate to the translational phase: a wing can travel at high angles of attack for a brief period, generating extra lift with a large leading-edge vortex, before it stalls. For the rotational mechanisms, a leading-edge vortex is created as the wing flips over in preparation for the subsequent half-stroke. The wing is thought to recapture this vorticity upon translation and thus to enhance its circulation. In principle, the pattern of vortex shedding should readily distinguish between the two classes of mechanisms: if the leading-edge vortex is created during the course of downstroke, then dynamic stall is operating; if the vortex is created as the wing flips over during pronation but stays attached on the downstroke, then rotational mechanisms are implicated.

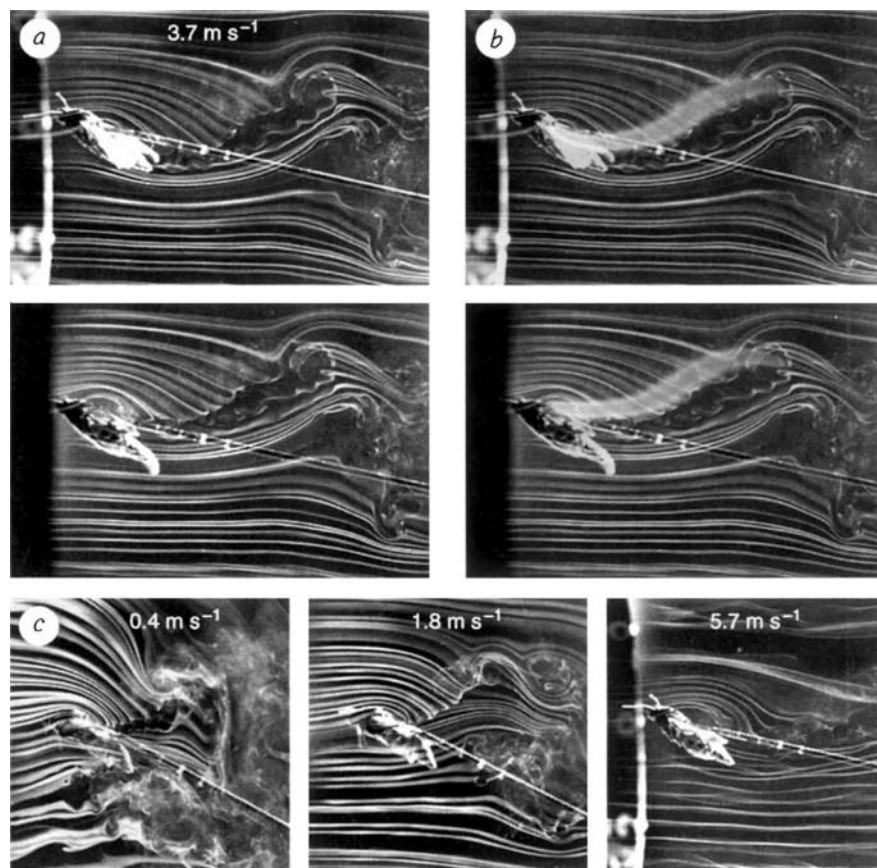


FIG. 1 Flow visualization around a female hawkmoth (mass 1.83 g) late in the downstroke. a, Stereo pair at 3.7 m s^{-1} ; b, stereo drawing superimposed on a to trace the main vortex structures; the top and bottom images correspond to the right- and left-hand views, respectively, in this stereo layout. c, Changes in the size of the leading-edge vortex at other speeds (0.4 , 1.8 and 5.7 m s^{-1}).

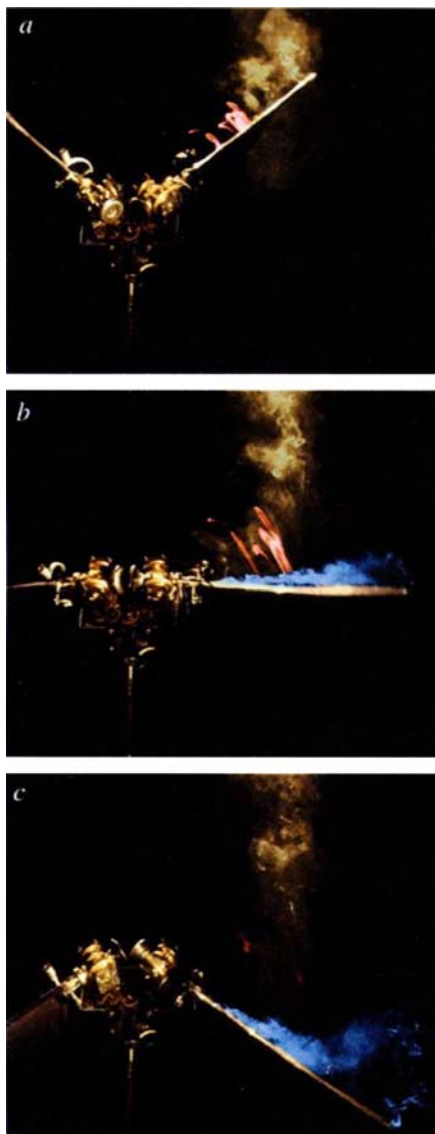


FIG. 2 Flapper flow visualization with smoke released from the leading edge over the inner half of one wing. The camera view is from above and in front of the 'hovering' flapper, parallel to the wing surfaces on the downstroke. *a*, At the end of pronation; *b*, just after the middle of the downstroke; *c*, near the end of the downstroke. The leading-edge vortex formed during pronation (red) is not recaptured by the wing, and a new leading-edge vortex (blue) is created during the translational phase of the downstroke. During the latter half of the downstroke, the core of the leading-edge vortex breaks down at 60–70% of the wing length and separates from the wing.

Unfortunately, the flow visualization results for tethered insects are not so clear-cut because of the difficulty in obtaining good flow visualization near the wings. To investigate the flow in more detail, we therefore built a three-dimensional mechanical model that closely mimics the wing movements of a hovering *Manduca*. This 'flapper' facilitated visualization by releasing smoke from the leading edge of the wing right into the core of the vortex (for further details, see refs 26, 27).

The model, with a 1.03-m wingspan, is about ten times larger than *Manduca*. To preserve aerodynamic similarity, a wingbeat frequency of 0.3 Hz was used to keep the Reynolds number constant (the ratio of inertial to viscous forces in the air): the frequency of *Manduca* is about 26 Hz. Another similarity condi-

tion that might prove important was also met. During a downstroke, the high angle of attack of the wings might cause free vortex shedding, which occurs at a frequency determined by the Strouhal number. For a hovering model at a given Reynolds number, the frequency of free vortex shedding is proportional to the wingbeat frequency. Thus the timing of any free vortex shedding is not altered relative to the forced vortex shedding imposed by the wingbeat. Any second-order patterns of vortex shedding for the flapper should therefore be similar to those for *Manduca*.

Flow visualization results with the 'hovering' flapper clearly identify dynamic stall as the high-lift mechanism. Figure 2 shows a downstroke, filmed obliquely from above and in front. Smoke was released from the leading edge over the inner half of one wing. Figure 2*a* is at the end of pronation, the rotation that precedes the downstroke; the leading-edge vortex formed as the air curled around the leading edge is coloured red, and the smoke remaining from the upstroke is uncoloured. The leading-edge vortex of pronation is quite two-dimensional in structure, with the swirls confined to planes perpendicular to the wing length. During the course of the downstroke, however, this vortex is left behind (Fig. 2*b*, *c*); the leading-edge vortex of pronation is not recaptured by the wing. A new leading-edge vortex, coloured blue, is instead created by the translational motion of the downstroke.

This leading-edge vortex appears at the beginning of the downstroke and persists until the end. As with the *Manduca* results, it grows larger during the course of the downstroke and towards the wingtip. Figure 3 shows the diameter and location of the vortex, as measured in light-slice experiments, for a time corresponding to that in Fig. 2*b*. The leading-edge vortex is a conical spiral, enlarging as it is swept along the wing by an axial (spanwise) flow. Under favourable conditions, the helix angle of clearly defined streaklines in the vortex could be measured. From photographs, the helix angle was estimated to be 46° (s.e.m. = 3° ; $n = 15$) with no significant spanwise changes. The magnitude of the axial flow was measured by manually releasing 'blobs' of smoke near the wing base and tracking their positions in video sequences. The axial velocity V_a increased steadily along the inner half of the wing, but then declined and became more variable (Fig. 4). The magnitude of the axial velocities was surprisingly large: near the maximum, values of V_a were fully comparable with the mean wingtip velocity, about 0.5 m s^{-1} .

The downstroke leading-edge vortex is stable and remains attached along the wing until just after the middle of the downstroke (Fig. 2*b*). The vortex core then appears to break down at about 60–70% of the wing length, and the tip region separates from the wing during the latter half of the downstroke (Fig. 2*c*). The variability in V_a (Fig. 4) over this wing region is probably attributable to the vortex breakdown and separation. The separated vortex feeds into a large tip vortex, which also agrees with the *Manduca* visualization results. The cause of vortex core breakdown is unclear, but it corresponds temporally with the onset of wing deceleration after the middle of the downstroke. Spatially, it coincides with the point where the leading edge begins to rake back in *Manduca*, and where the trailing edge bows out because of the hindwing: these planform characteristics can be seen in Fig. 3.

The swirl velocity of the leading-edge vortex could not be measured in light-slice experiments, because the smoke blobs moved out of the slices too quickly. However, the swirl velocity must be equal to the axial velocity because the helix pitch angle is almost 45° . By combining this value with measurements of the vortex diameter, as in Fig. 3, the circulation Γ_{le} of the leading-edge vortex can be calculated. Results are shown in Fig. 5 for the early, middle and late downstroke; the early stage is soon after Fig. 2*a*, whereas the middle and late stages roughly correspond to Fig. 2*b* and *c*, respectively. The circulation clearly increases during the first half of the downstroke, which is consistent with dynamic stall but not with the recaptured vortex of rotational lift mechanisms. The circulation decreases over the outer wing region in the late

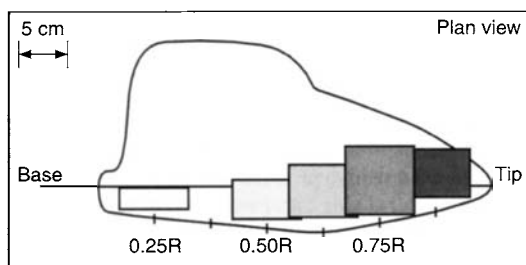


FIG. 3 Planform view of the wing, showing sections of the leading-edge vortex at five spanwise locations (as fractions of the winglength R) just after the middle of the downstroke, as in Fig. 2b. Sections were illuminated by a 7-cm-wide light slice perpendicular to the wing axis; the dimensions of each pattern show the width of the slice, the vortex diameter and its position on the planform. The leading-edge vortex is a conical spiral, enlarging as it is swept along the wing by the axial flow. The vortex separates from the wing near the wingtip, joining up with the tip vortex, and at $0.87R$ it has lifted away from the wing surface.

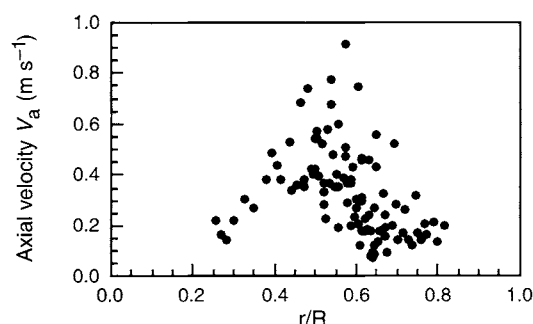


FIG. 4 Axial (spanwise) velocity V_a in the leading-edge vortex, estimated by tracking 'blobs' of smoke released from the wing base. The blobs reached different radial positions r (expressed as a fraction of the winglength R) at various stages of the downstroke; results are pooled for the whole downstroke.

downstroke, a result of breakdown and separation of the leading-edge vortex.

The spanwise variation of circulation is also consistent with dynamic stall but not the rotational mechanisms. For dynamic stall, the circulation should be proportional to the product rc , where r is radial position and c is the local chord; for rotational mechanisms, circulation should be proportional to c^2 (ref. 15). These two alternatives are shown in Fig. 5, scaled to give the same lift in the middle of the downstroke. The measured circulation for the middle of the downstroke agrees very well with dynamic stall, and the large circulation over the inner half of the wing that would result from rotational mechanisms is absent.

Direct observation of the patterns of vortex shedding, the growth of circulation early in the downstroke, and the spanwise variation in circulation all point unequivocally to dynamic stall as the aerodynamic mechanism responsible for the downstroke leading-edge vortex. Is this vortex strong enough to explain the high lift required for hovering *Manduca*? The wake of the flapper has been analysed in a separate study²⁶, and the momentum imparted to the air corresponds to a mean lift force on the downstroke of about 1.5 times the weight of *Manduca*. The scaling of winglength and frequency for the flapper will leave the aerodynamic forces unchanged, so this is the same force that *Manduca* should generate. The magnitude of the downstroke lift is, therefore, more than adequate.

Lift enhancement by the spiral leading-edge vortex bears several similarities to the high-lift devices employed on certain man-made wings. The potential of 'attached' vortices to augment

lift has long been recognized in aerodynamics, but an axial (spanwise) flow component is essential for the stability of such vortices (reviewed in ref. 28). By convecting the vorticity out to the wingtip, this flow prevents it from accumulating into a large vortex that would be unstable under two-dimensional conditions. The axial flow can be induced by active spanwise suction or blowing over the upper wing surface; for delta wings, it is created instead by the flow component parallel to the swept leading edge. The conical, spiral vortex of the flapper is, in fact, remarkably similar in form to that over delta wings²⁷. However, the mechanism of vortex generation for stationary, swept delta wings is quite different from that for hovering insects with flapping, non-swept wings.

Axial flow in the leading-edge vortex has not been observed in previous two-dimensional experiments relevant to insect flight, but the spanwise pressure gradient necessary to drive the axial flow is, by definition, absent in two-dimensional studies. Maxworth's¹⁰ three-dimensional model of the specialized fling motion is the only other case where axial flow has been reported, and it now seems likely that this three-dimensional flow pattern is a common feature of insect flight. It seems reasonable to assume that the flow is generated either by the dynamic pressure gradient associated with the velocity gradient along the flapping wing, by 'centrifugal' acceleration in the boundary layer, or by the induced velocity field of the spiral vortex lines. Helicopter rotors and windturbine blades also experience such spanwise pressure gradients and centrifugal accelerations, but large-scale spanwise flows have not been observed²⁹; however, there is some evidence that a spanwise flow influences their stall characteristics³⁰. It may be that the Reynolds numbers are too high for a large leading-edge vortex to persist, or that the spanwise flow component is reduced for high-aspect-ratio airfoils like helicopter blades. The exact conditions for establishing axial flow in a leading-edge vortex for rotary wings are not yet understood. □

Methods

Hawkmoth visualization. Moths were tethered 5 cm downstream from a 26-cm-diameter open jet wind-tunnel. The body orientation was set to that recorded by high-speed video (Kodak Ektapro 1000, operating at 1,000 frames per second) during free, feeding flights in the wind-tunnel at the same speeds. The tether was mounted on a balance, and experiments were rejected if the moth did not support at least 70% of its weight. Smoke was produced by an FVSP/E smoke generator (Nutem) with medicinal white oil (Shell Ondina EL). Stereo photographs were recorded on two Nikon F3 cameras with optical axes converging by 10° . Illumination was provided by four strobe lamps, and the cameras were set at f2.8 and 1/30 s. Ilford HP5 Plus film was processed at 1600 ASA with Microphen developer.

Flapper. The flapper body houses four servo motors and an elaborate gearbox to drive the wing movements, which are mechanically coupled for the left and right wings. As the hawkmoth is a functionally two-winged insect, the fore- and hindwing are constructed as one wing. The wings (2.3 mm thick) consist of a venation-like framework of brass tubes, with rigid and flexible joints as appropriate, covered on both sides with black, elastic cloth. A stiff brass smoke

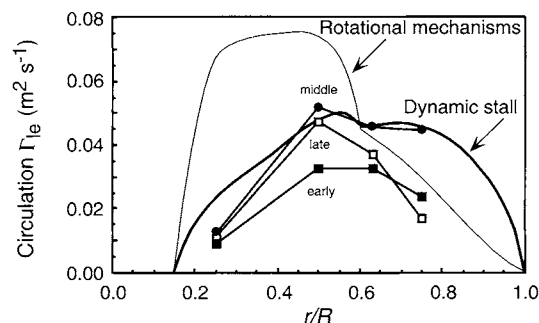


FIG. 5 Estimates of the circulation Γ_{le} of the leading-edge vortex at four spanwise positions for the early, middle and late downstroke. Circulation distributions corresponding to dynamic stall and the rotational mechanisms are also drawn, scaled to give the same lift as Γ_{le} at the middle of the downstroke.

rake extends along the leading edge, and a flexible wire along the trailing edge. Two servos set the angles of attack for the relatively stiff leading edge/wingtip region and for the more flexible trailing edge region. The other two servos flap the wings about their bases, and elevate them perpendicular to the flapping plane. Wing movements were controlled by a Macintosh Quadra 650 computer with D-A converters (NB-A06, National Instruments); custom software was written in LabView 3.0. Kinematic data for a hovering *Manduca* were taken from A.P.W. and C.P.E., manuscript in preparation. Photographic images were scanned and the smoke was then coloured in Photoshop 3.0.

Circulation estimates. Γ_{le} was calculated as $\pi d V_{\theta}$, where d is the diameter and V_{θ} is the swirl velocity of the leading-edge vortex. Because the helix-pitch angle is about 45° , V_{θ} is equal to the axial velocity V_a . Mean values of V_a were calculated for four spanwise positions from the data in Fig. 4. Circulation distributions in Fig. 5 were scaled to give equal lift in the middle of the downstroke by equating integrals of the product of circulation and flapping velocity.

Received 26 June; accepted 21 October 1996.

1. Weis-Fogh, T. *J. Exp. Biol.* **59**, 169–230 (1973).
2. Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 145–181 (1984).
3. Ennos, A. R. *J. Exp. Biol.* **142**, 49–85 (1989).
4. Dudley, R. & Ellington, C. P. *J. Exp. Biol.* **148**, 53–88 (1990).
5. Dudley, R. *J. Exp. Biol.* **198**, 1065–1070 (1995).
6. Cloupeau, M., Devillers, J. F. & Deveziaux, D. *J. Exp. Biol.* **80**, 1–15 (1979).
7. Wilkin, P. J. *J. Kansas Entomol. Soc.* **63**, 316–328 (1990).
8. Zanker, J. M. & Götz, K. G. *Phil. Trans. R. Soc. Lond. B* **327**, 19–44 (1990).
9. Wilkin, P. J. & Williams, M. H. *Physiol. Zool.* **66**, 1015–1044 (1993).
10. Maxworthy, T. *J. Fluid Mech.* **93**, 47–63 (1979).
11. Edwards, R. H. & Cheng, H. K. *J. Fluid Mech.* **120**, 463–473 (1982).
12. Spedding, G. R. & Maxworthy, T. *J. Fluid Mech.* **165**, 247–272 (1986).
13. Sunada, S., Kawachi, K., Watanabe, I. & Azuma, A. *J. exp. Biol.* **183**, 217–248 (1993).
14. Nachtigall, W. *J. Comp. Physiol.* **133**, 351–355 (1979).
15. Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 79–113 (1984).
16. Ellington, C. P. in *Biological Fluid Dynamics* (eds Ellington, C. P. & Pedley, T. J.) *Symp. Soc. Exp. Biol.* **49**, 109–129 (1995).
17. Brodsky, A. K. & Ivanov, V. D. *Zool. Zhurn.* **63**, 197–208 (1984).
18. Brodsky, A. K. & Grodnitsky, D. L. *Ent. Obozr.* **64**, 484–492 (1985).
19. Brodsky, A. K. *J. Exp. Biol.* **161**, 77–95 (1991).
20. Grodnitsky, D. L. & Morozov, P. P. *J. Exp. Biol.* **169**, 143–163 (1992).
21. Grodnitsky, D. L. & Morozov, P. P. *J. Exp. Biol.* **182**, 11–40 (1993).
22. Brodsky, A. K. *The Evolution of Insect Flight* (Oxford Univ. Press, Oxford, 1994).
23. Willmott, A. P., Ellington, C. P. & Thomas, A. L. R. *Phil. Trans. R. Soc. Lond. B* (in the press).
24. Dickinson, M. H. & Götz, K. G. *J. Exp. Biol.* **174**, 45–64 (1993).
25. Dickinson, M. H. *J. Exp. Biol.* **192**, 179–206 (1994).
26. Van den Berg, C. & Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* (in the press).
27. Van den Berg, C. & Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* (in the press).
28. Wu, J. Z., Vakili, A. D. & Wu, J. M. *Prog. Aerospace Sci.* **28**, 73–131 (1991).
29. De Vries, O. *Annu. Rev. Fluid Mech.* **15**, 77–96 (1983).
30. Harris, F. D. *J. Am. Helicopter Soc.* **11**, 1–21 (1966).

ACKNOWLEDGEMENTS. This work was supported by the Science and Engineering Research Council, the Hasselblad Foundation and an EC Human Capital and Mobility grant. We thank R. Holder for building the flapper and R. J. Wootton for designing the wings.

CORRESPONDENCE and requests for materials should be addressed to C.P.E. (e-mail: cpe10@cam.ac.uk).

Altered segmental identity and abnormal migration of motor neurons in mice lacking *Hoxb-1*

Michèle Studer, Andrew Lumsden*,
Linda Ariza-McNaughton, Allan Bradley†
& Robb Krumlauf

Division of Developmental Neurobiology, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

* Department of Developmental Neurobiology, UMDS, Guy's Hospital, London SE1 9RT, UK

† Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

SEGMENTATION of the vertebrate hindbrain into rhombomeres is important for the anterior–posterior arrangement of cranial motor nuclei and efferent nerves¹. Underlying this reiterated organization, *Hox* genes display segmentally restricted domains of expression^{2–4}, such as expression of *Hoxb-1* (refs 5, 6) in

rhombomere 4 (r4). Here we report that absence of *Hoxb-1* leads to changes in r4 identity. In mutant mouse embryos, molecular markers indicate that patterning of r4 is initiated properly but not maintained. Cellular analysis by DiI tracing reveals that the r4-specific facial branchiomotor (FBM) and contralateral vestibuloacoustic efferent (CVA) neurons are incorrectly specified. In wild-type mice CVA neurons migrate from r4 into the contralateral side⁷, and we found in lineage analysis that FBM neurons migrate from r4 into r5. In mutants, motor neurons differentiate but the CVA and FBM neurons fail to migrate into their proper positions. Instead, they form a motor nucleus which migrates atypically, and there is a subsequent loss of the facial motor nerve. These results demonstrate that, as a part of its role in maintaining rhombomere identity, *Hoxb-1* is involved in controlling migratory properties of motor neurons in the hindbrain.

Despite numerous studies in which targeted mutations of *Hox* genes have been generated³, their functional role in hindbrain development remains unclear. Loss of function of *Hoxa-1* results in the deletion of rhombomeres^{8–10}, suggesting that it controls segmentation, whereas overexpression of *Hoxa-1* (refs 11, 12) or *Hoxb-1* (refs 11–13) induces changes in patterning, which is more consistent with an involvement in conferring rhombomere identity. Because the r4 expression of *Hoxb-1* is unique amongst the *labial* group (reviewed in ref. 4), removal of its function might help to distinguish between these alternatives. Accordingly, we generated two loss-of-function mutations in the *Hoxb-1* locus by homologous recombination in embryonic stem cells (Fig. 1a, b). For both mutant alleles, heterozygous mice were indistinguishable from wild-type animals, but 98% of homozygotes died within 24 h of birth. We used a polyclonal antibody raised against *Hoxb-1* (ref. 13) to confirm the absence of *Hoxb-1* in homozygous mutant embryos (Fig. 1c).

Our previous studies indicated that the r4 expression of *Hoxb-1* is dependent upon an enhancer with a binding site for labial-related proteins, suggesting that r4 is controlled by an auto- or cross-regulatory loop¹³. Because the r4 expression is not lost in *Hoxa-1* mutants^{8–10}, we investigated whether *Hoxb-1* itself is required by examining the consequences of loss of *Hoxb-1* protein on its own regulation. Transgenic lines carrying an alkaline phosphatase reporter construct (HPAP) under the control of the *Hoxb-1* r4 regulatory region¹⁴ were assayed for reporter expression at 9.5 days post coitum (d.p.c.) in both mutant and wild-type backgrounds (Fig. 1d–g). Absence of *Hoxb-1* results in the loss of reporter expression specifically in r4, leaving the limb bud staining unaffected (Fig. 1f, g). Thus r4 expression of *Hoxb-1* is strictly dependent on its own product and confirms that direct auto-regulatory mechanisms¹³ are an important component of segmental expression.

To investigate whether hindbrain segmentation is affected in *Hoxb-1*^{−/−} mice, we analysed the pattern of segmentally expressed markers. The expression of *Krox20* (ref. 15) and *Sek1* (ref. 16) in r3 and r5, of *kreisler* (ref. 17) in r5 and r6, and of *Sek2* (refs. 18, 19) in r4 was not altered in early stages (Fig. 2a–f; data not shown), and the relative size of the rhombomeres appeared normal. Therefore, the initial sequence of events leading to the generation and molecular patterning of r4 and neighbouring rhombomeres does not depend on *Hoxb-1*.

In contrast, markers expressed in r4 from 8.25 d.p.c. onwards, a period coinciding with the upregulation of *Hoxb-1* in r4, did show alterations. In 8.5 d.p.c. embryos (7–9 somites) mutants have no expression of *Wnt8* (ref. 20) in r4 (Fig. 2g, h). The antisense strand of *Hoxb-3* is expressed in r3 and more posteriorly, with high levels in r4 and r6 (Fig. 2i)²¹, and *CRABPI* (refs 22, 23) is expressed at low levels in r2 and r3, with high levels in r4 to r6 (Fig. 3k). In homozygous mutant embryos, both these probes show a specific lack of upregulation in the r4 domain, although other regions are unaltered (Fig. 2j, l).

We then examined whether these alterations reflect a change in segmental identity. To assay for r2 characteristics, we generated a transgenic line containing an HPAP reporter gene under control

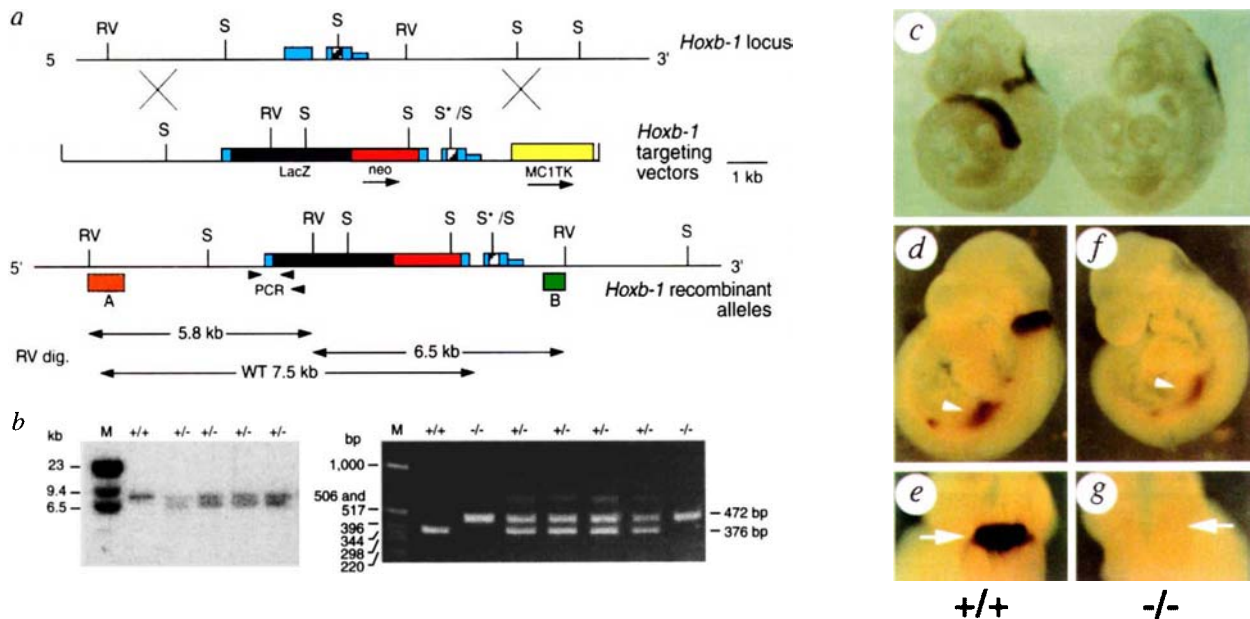


FIG. 1 Targeted inactivation of *Hoxb-1* and abolishment of r4 autoregulation. *a*, Structure of the mouse *Hoxb-1* locus, replacement targeting vectors, and recombinant alleles. S*/S indicates that one targeting vector also had a frameshift in the *SstI* site of the homeobox. Boxes A and B correspond to probes. Arrowheads indicate the position of PCR primers used in genotyping. *b*, Embryonic stem cell DNA blot probed with B (left) and PCR products from 9.5-d.p.c. yolk-sac DNA (right). *c*, Whole-mount immunocytochemistry with anti-*Hoxb-1* antibody¹³ with wild-type (left) and homozygous mutant (right) embryos at 9.5 d.p.c. *d*, *e*, Lateral

and dorsal views of a 9.5-d.p.c. wild-type embryo for *Hoxb-1* carrying a transgene with the *Hoxb-1* r4 enhancer^{26,34} linked to a human alkaline phosphatase (HPAP) reporter gene¹⁴. *f*, *g*, Lateral and dorsal views of a 9.5-d.p.c. *Hoxb-1*^{-/-} embryo carrying the same r4 transgenic reporter. Arrows indicate position of r4, and arrowheads indicate reporter staining in limb-bud mesenchyme. Note the specific loss of r4 expression in mutants. Wild-type (+/+) and homozygous (-/-) mutant genotypes are indicated. S, *SstI* S* mutant *SstI*; RV, *EcoRV*.

of an r2 enhancer²⁴, and used anti-Sek1 antibodies²⁵. In wild-type embryos, *Sek1* is expressed uniformly in r3 and r5 at high levels and dorsally in r2 at lower levels²⁵, but is not expressed in r4 (Fig. 2m). In *Hoxb-1*^{-/-} mice, ectopic expression identical to that found in r2 is present in r4 (Fig. 2n). In addition, when any of three different r2-HPAP reporter lines are mated into the *Hoxb-1* mutant background, ectopic patches of staining are induced in r4 (Fig. 2o, p). Although the induction of the r2 reporter is only seen in a subset of r4 cells, this occurs with 100% penetrance in the mutant embryos at all stages examined after 9.0 d.p.c., and is never observed in any of the three parental r2-HPAP lines. This suggests that additional components may be involved either in maintaining r4 identity or in generating full r2 identity. Together, these molecular data indicate that in *Hoxb-1*^{-/-} embryos the early patterning of r4 is initiated properly but is not maintained, leading to a partial transformation or intermediate identity.

Segmentation in the hindbrain is correlated with the reiterated organization of the cranial sensory ganglia and the branchial and visceral motor nerves¹. To determine whether there were anatomical changes in the r4 region of mutant embryos, we analysed axonal projections both in the peripheral (PNS) and central (CNS) nervous systems. In the PNS, markers revealed no abnormalities in the morphology of sensory ganglia and their axonal projections (data not shown). In the murine CNS, the VIIth cranial nerve (facial and intermediate) exits r4 but collects axons from three distinct subpopulations of neurons in r4 and r5: facial branchiomotor neurons (FBM)²⁶, contralateral vestibuloacoustic efferent neurons (CVA)⁷, and visceromotor neurons (VM) of the superior salivary nucleus²⁷. To visualize the positions and trajectories of these efferent neurons we used retrograde DiI tracing in embryos at 10.5 to 11.5 d.p.c. We found that two subpopulations, FBM and CVA, were missing from their normal location in the mutants, suggesting that there may be defects in either migration or axon guidance (Fig. 3a-d). These are precisely the two neuronal populations which colocalize with high levels of *Hoxb-1* expression at these stages²⁶. The third subpopulation, VM

neurons, differentiate in r5 and send projections laterally and anteriorly, exiting r4 with the VIIth (intermediate) nerve²⁷ (Fig. 3e). By using anterograde DiI tracing from the r4 and r5 basal plate, we found that in mutant embryos VM cells had abnormal axonal trajectories that converged to a new ectopic exit point in r5 (Fig. 3f). Because *Hoxb-1* is not expressed in these cells their altered r5 projections are likely to be a secondary consequence of the defects found in r4, illustrating the importance of signalling between rhombomeres for proper neuronal patterning. Thus all three subpopulations of the facial motor nerve are affected in the mutants.

To understand the absence of the FBM and CVA neurons, and distinguish between migratory and guidance defects, we needed to determine the origin of these two populations. The r4-specific CVA neurons normally migrate medially across the floor plate into the contralateral r4 (ref. 7); however, the origin of the FBM neurons along the border of the floor plate in r5 is not clear. Therefore, we developed a hindbrain explant culture system using a *Hoxb-1/lacZ* transgenic line²⁶ for lineage analysis. Explants from embryos at 9.75 d.p.c., stained with a viable fluorescent β -galactosidase substrate, showed reporter expression only in r4, which facilitated the focal DiI labelling of r4-specific cells (Fig. 3g, h). After 26 h of culture, reporter staining and r4 labelled cells were found ventrally in r5 (Fig. 3i, j). In the explants, cell bodies that migrated posteriorly along the midline had axonal projections extending into r4 (Fig. 3k, l). Labelled cells are not detected in r3 or other regions of r5, showing that the r4-derived FBM neurons migrate into r5 only along the midline.

The absence of DiI retrograde labelling of r5 FBM neurons could arise from a failure of these cells to send projections to the periphery, although the cell bodies remain in their expected locations. Therefore, we used *c-ret* and *Islet1* as motor-neuron markers to assay independently for the presence or absence of motor cell bodies. At 12.5 d.p.c., the normal *Islet1* (ref. 28) and *c-ret*²⁹ expression in FBM neurons along the midline is not detected in serial sections through the mutants (Fig. 4a, b; data

not shown). This shows that FBM cell bodies are not positioned normally, and combined with the failure to detect extensive abnormal projections in our anterograde tracing from the basal plate of r4 or r5 (Fig. 3f), argue for a defect in migration rather than guidance. At 12.5 d.p.c., *Islet1* also marks a distinctive posterior and lateral migration of the facial motor nerve in wild-type embryos²⁷ that is not detected in the mutants (Fig. 4a, b).

However, there was ectopic staining adjacent to the exit point of the VIIth nerve, which suggests that facial motor neurons had undergone a lateral migration more similar to those of the trigeminal nucleus (Figs 4a–d and 5b). We examined the fate of this ectopic neuronal subpopulation in embryos at 15.5 to 16.5 d.p.c., and found in serial sections that *Islet1* and *c-ret* expression were greatly diminished and then lost in mutant embryos (Fig. 4e–h);

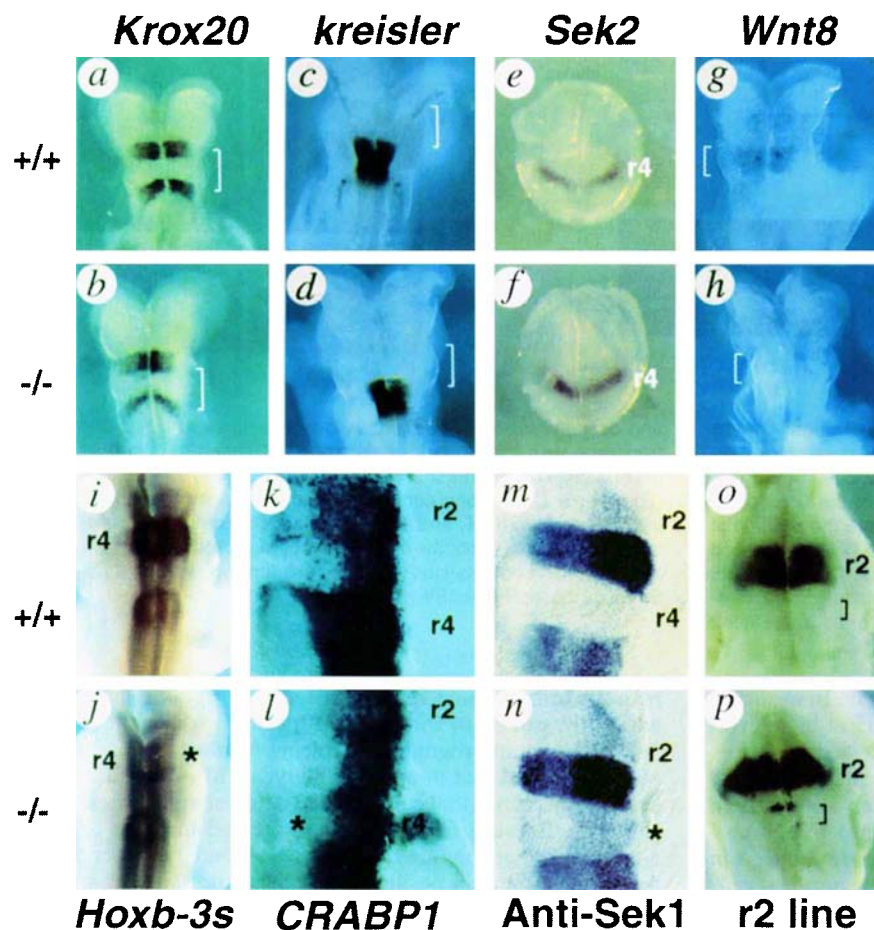


FIG. 2 Changes of r4 identity in *Hoxb-1* mutant embryos. Dorsal views of embryos probed with a, b, *Krox20* for r3 and r5 (8.5 d.p.c.); c, d, *kreisler* for r5 and r6 (8.75 d.p.c.); e, f, *Sek2* for r4 (8.0 d.p.c.); g, h, *Wnt8* for r4 (8.5 d.p.c.); i, j, *Hoxb-3s* for r4 and r6 (9.5 d.p.c.); k, l, *CRABP1* for r2 to r5 (9.5 d.p.c.); m, n, anti-*Sek1* for r2, r3 and r5 (9.5 d.p.c.); and o, p, HPAP transgenic reporter for r2 (9.5 d.p.c.). Brackets indicate presumptive r4 region; asterisk indicates changes in r4 expression; k–n, flat mounts of the hindbrain. (+/+) and (–/–) indicate wild-type and homozygous mutant embryos, respectively.

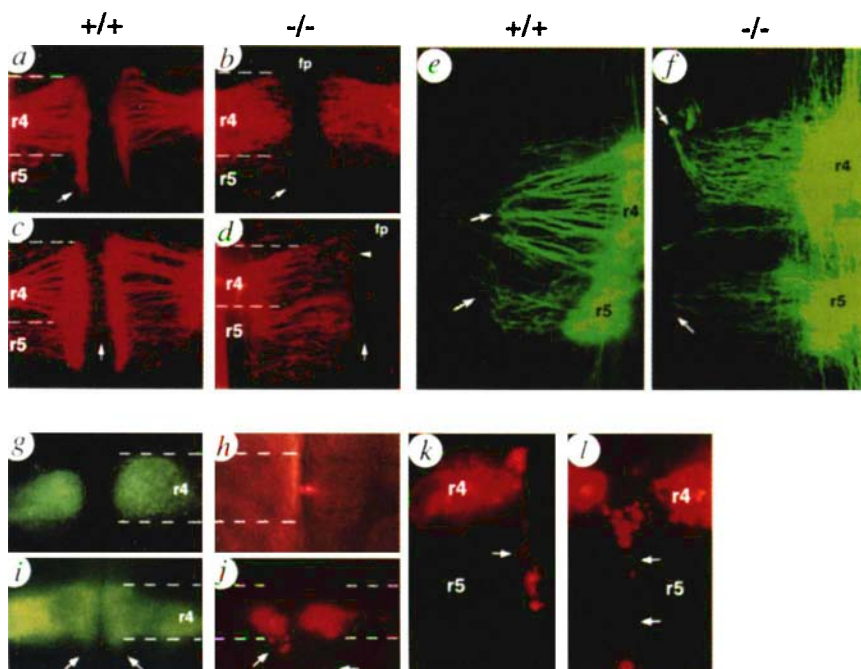


FIG. 3 Analysis of migration and projections of the facial motor neurons in the hindbrain. a, b, Retrograde Dil labelling of the facial motor nerve in 10.5-d.p.c. wild-type (a) and mutant (b) embryos. Arrows indicate FBM neurons that are missing in the mutants. c, d, Retrograde labelling in 11.5-d.p.c. wild-type (c) and mutant (d) embryos. Arrows indicate CVA efferent neurons crossing the floor plate that are also missing in the mutants; arrowhead indicates dispersed cell bodies in r4. e, f, Anterograde Dil labelling of r4 and r5 motor components in 10.5-d.p.c. wild-type (e) and mutant (f) embryos. In wild type, labelled axons from r5 project anterior to an exit point in r4, but in mutants projections do not turn and may exit r5. g, i, *Hoxb-1* r4–*lacZ* reporter expression²⁶ assayed with viable fluorescent β -galactosidase substrate in 9.75-d.p.c. hind-brain explants cultured for 2 h (g) or 26 h (i). Arrows indicate staining in r5. h–l, Dil labelling of r4 cells in explant after 2 h culture (h), and presence of labelled cells in r5 after 26 h culture (j–l). Arrows indicate Dil-labelled FBM cell bodies in r5 that extend axons into r4. Wild-type (+/+) and *Hoxb-1* mutant (–/–) embryos are indicated.

data not shown).

Together these results show, by two independent routes (retrograde labelling and motor-neuron markers), that facial neurons and peripheral projections are initially formed in the mutant r4. However, the neurons have altered migratory behaviours (summarized in Fig. 5), and fail to undergo the normal active movement of cell bodies demonstrated by the explant lineage analysis after they have extended their axons outside the brain. The subsequent loss of the facial motor nucleus suggests that these neurons are no longer able to receive or interpret signals necessary for their survival.

We have shown that *Hoxb-1* is important for maintaining and/or specifying r4 identity. Loss of *Hoxb-1* function, unlike that of its labial group paralogue *Hoxa-1* (refs 8–10), does not result in rhombomere deletions, demonstrating that it is not required for generating hindbrain segments. The failure of r4 markers to be upregulated (*Wnt8*, *Hoxb3s*, *CRABP1* and *Hoxb-1*) and the associated ectopic expression of r2 markers (*Sek1* and r2-HPAP) suggest that r4 adopts an altered identity. Consistent with this, in *Hoxb-1* mutants the posterior migratory behaviour of the r4 neurons is abnormal, and they adopt a lateral movement typical of the trigeminal neurons in r2 (Fig. 5b)³⁰. The failure of the trans-midline migration of CVA neurons also confirms the loss of

specific migratory capability by r4 neurons in the mutants. In conclusion, these alterations strongly suggest that *Hoxb-1* is involved in the regulation of neuronal migration in the hindbrain, as part of its role in maintaining segmental identity. □

Methods

Targeted disruption of *Hoxb-1* gene and generation of transgenic lines. *KpnI* linearized constructs (Fig. 1a) and a positive/negative selection targeting strategy were used in AB2.2 embryonic stem cells³¹. The *SstI** site was destroyed by cutting with *SstI* and blunt-ending. Independent targeted clones with either the *SstI* or *SstI** alleles produced germline chimaeras and gave identical results. Phenotypes were analysed in embryos derived from intercrossing heterozygotes from a mixed 129/Sv and C57/B6 background. Primer sequences: wild-type primer a, agcttcagctctgtgacatavtgc, and primer b, gaataatactgagaagcccatagctgg; mutant primer c, tagatggcgcatcgtaaccgtgcat. Three r2-HPAP transgenic lines were produced using a 1.1-kb fragment of *Hoxa-2* (ref. 24) in antisense orientation linked to a human placental alkaline phosphatase reporter gene (HPAP).

In situ hybridization and immunocytochemistry. Whole-mount *in situ* hybridization³² was performed with digoxigenin-labelled probes. Whole-mount immunocytochemistry was performed with anti-*Hoxb-1* and anti-*Sek1* antibodies²⁵, except that embryos for *Hoxb-1* were fixed in 3.7% formaldehyde (A. Gould, unpublished data). Serial cryostat sections 20 µm thick from fresh frozen embryos were hybridized and washed³³.

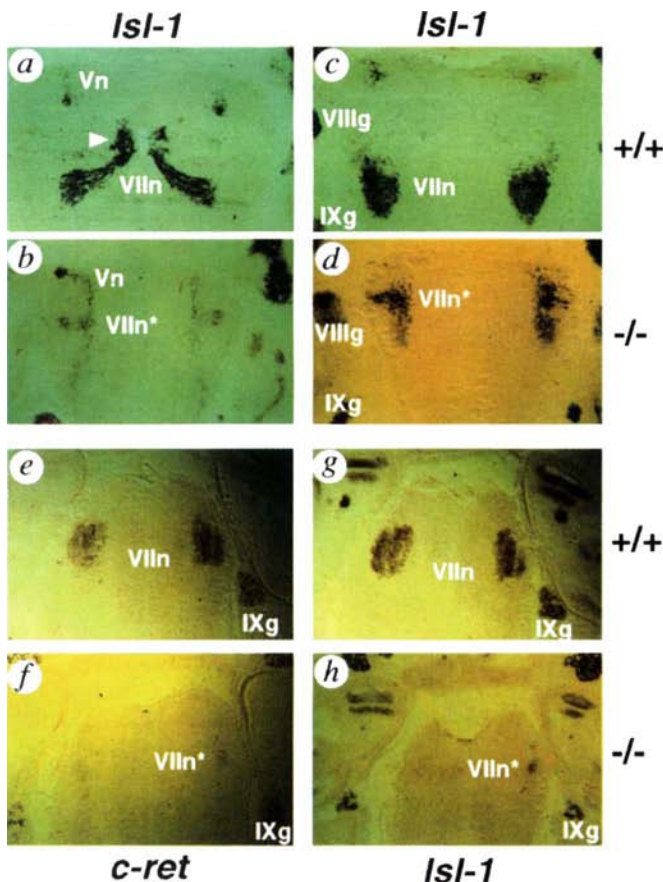


FIG 4 Analysis of motor-neuron markers. a–d, Near-adjacent coronal sections through the hindbrain of 12.5-d.p.c. wild-type (a, c) and mutant (b, d) embryos probed with *Isl-1*. Arrowhead in a marks midline FBM neurons, missing in mutants. Note also that misspecified facial motor neurons (Vlln*) do not undergo a normal caudal migration (compare a and b) and become situated more rostrally in register with the Vllg (compare c and d). e–h, Near-adjacent coronal sections through the medulla oblongata of 15.5-d.p.c. wild-type (e, g) and mutant (f, h) embryos. The sections in f and h are the only sections to show weak positive staining of the facial nucleus (Vlln*) indicating that it is almost completely lost by this stage. Vn, trigeminal motor nucleus; Vlln, facial motor nucleus; Vllg, vestibulo-cochlear ganglion; IXg, glossopharyngeal ganglion.

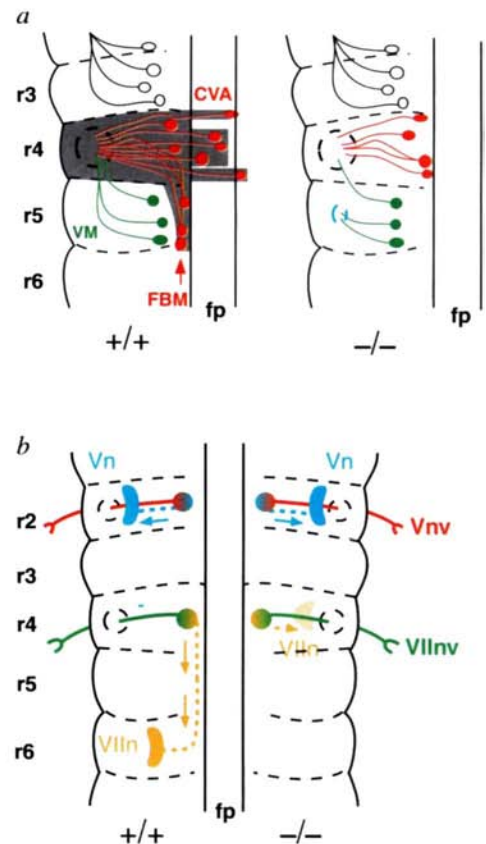


FIG 5 Summary diagrams of facial motor-neuron patterning. Wild-type (+/+) and *Hoxb-1* mutant (-/-) embryos. a, Organization of the facial motor nerve in the hindbrain. In mutants the CVA and FBM neurons (red) fail to migrate from r4, and the VM (green) have an ectopic exit point in r5 (blue). The grey area represents overlap of *Hoxb-1* expression with FBM and CVA neurons. b, Migratory trajectories of the trigeminal (V) and facial (VII) motor nuclei. In wild-type embryos, after the initial projections of axons to their exit points (red, trigeminal; green, facial), the nuclei condense and migrate either laterally towards (blue, trigeminal) or caudally away from (yellow, facial) their exit points. In (-/-) mutants the facial nucleus migrates laterally, like the trigeminal.

Retrograde and anterograde labelling. Embryos were fixed in 4% paraformaldehyde in phosphate buffered saline (PBS) and injected with the carbocyanine tracer Dil. Retrograde labelling and flat mount preparations were as described²⁶. For anterograde labelling a series of small tracer injections were made adjacent to the floor plate along r4 and r5, and specimens were bisected along the midline and viewed under a confocal microscope.

Explant culture and Dil analysis. Hindbrain tissue (spanning from midbrain to somite 3/4) was removed from embryos at 9.75 d.p.c. opened along the dorsal midline and cultured at 37 °C on Millicell-CM filters (Millipore) in DMEM with 10% fetal calf serum. Reported staining in the explants was performed using the DetectaGene Green CMFDG lacZ gene expression kit (Molecular Probes). Focal injections of Dil (0.5% in ethanol) into small groups of cells were made by iontophoresis.

Received 23 September; accepted 25 November 1996.

1. Lumsden, A. & Keynes, R. *Nature* **337**, 424–428 (1989).
2. Hunt, P. et al. *Nature* **353**, 891–864 (1991).
3. Krumlauf, R. *Cell* **78**, 191–201 (1994).
4. Keynes, R. & Krumlauf, R. *Annu. Rev. Neurosci.* **17**, 109–132 (1994).
5. Murphy, P., Davidson, D. R. & Hill, R. E. *Nature* **341**, 156–159 (1989).
6. Wilkinson, D. G., Bhatt, S., Cook, M., Boncinelli, E. & Krumlauf, R. *Nature* **341**, 405–409 (1989).
7. Simon, H. & Lumsden, A. *Neuron* **11**, 209–220 (1993).
8. Carpenter, E. M., Goddard, J. M., Chisaka, O., Manley, N. R. & Capecchi, M. R. *Development* **118**, 1063–1075 (1993).
9. Mark, M. et al. *Development* **119**, 219–338 (1993).
10. Dollé, P. et al. *Proc. Natl Acad. Sci. USA* **90**, 7666–7670 (1993).
11. Alexandre, D. et al. *Development* **122**, 723–746 (1996).
12. Zhang, M. et al. *Development* **120**, 2431–2442 (1994).
13. Pöppert, H. et al. *Cell* **81**, 1031–1042 (1995).
14. Itasaki, N., Sharpe, J., Morrison, A. & Krumlauf, R. *Neuron* **16**, 487–500 (1996).
15. Wilkinson, D. G., Bhatt, S., Chavrier, P., Bravo, R. & Charnay, P. *Nature* **337**, 461–464 (1989).

16. Nieto, M. A., Gilardi-Hebenstreit, P., Charnay, P. & Wilkinson, D. *Development* **116**, 1137–1150 (1992).
17. Cordes, S. P. & Barsh, G. S. *Cell* **79**, 1025–1034 (1994).
18. Ruiz, J. & Robertson, D. *Mech. Dev.* **46**, 87–100 (1994).
19. Becker, N. et al. *Mech. Dev.* **47**, 3–18 (1994).
20. Hume, C. R. & Dodd, J. *Development* **119**, 1147–1160 (1993).
21. Swiatek, P. J. & Gridley, T. *Genes Dev.* **7**, 2071–2084 (1993).
22. Ruberte, E., Friederich, V., Moriss-Kay, G. & Chambon, P. *Development* **115**, 973–989 (1992).
23. Maden, M. et al. *Mech. Dev.* **37**, 13–23 (1992).
24. Frasch, M., Chen, X. & Lufkin, T. *Development* **121**, 957–974 (1995).
25. Irving, C., Nieto, M., Das Gupta, R., Charnay, P. & Wilkinson, D. *Dev. Biol.* **173**, 26–38 (1996).
26. Marshall, H. et al. *Nature* **360**, 737–741 (1992).
27. McKay, I., Lewis, J. & Lumsden, A. *Mol. Cell. Neurobiol.* (in the press).
28. Tsuchida, T. et al. *Cell* **79**, 957–970 (1994).
29. Pachnis, V., Maniok, B. & Costantini, F. *Development* **119**, 1005–1017 (1993).
30. Altman, J. & Bayer, S. *Development of the Cranial Nerve Ganglia and Related Nuclei in the Rat* (Springer, Berlin, 1982).
31. Ramirez-Solis, R., Davis, A. & Bradley, A. in *Guide to Techniques in Mouse Development* (eds Wasserman, P. & DePamphilis, M.) 885–878 (Academic, 1993).
32. Wilkinson, D. & Green, J. in *Postimplantation Mouse Embryos: a Practical Approach* (eds Copp, A. J., Cockcroft, D. L. & Hames, B. D.) 155–171 (IRL, Oxford, 1990).
33. Schaefer-Wiemers, N. & Gerfin-Moser, A. *Histochemistry* **100**, 431–440 (1993).
34. Studer, M., Pöppert, H., Marshall, H., Kuroiwa, A. & Krumlauf, R. *Science* **265**, 1728–1732 (1994).

ACKNOWLEDGEMENTS. We thank D. Wilkinson (Sek1, Krox20), P. Charnay (Sek2), S. Lee and A. McMahon (Wnt8), F. Rijli (CRABP1), V. Pachnis (c-ret, Islet1), A. Gould (Hoxb-1) and S. Cordes and G. Barsh (kreisler) for probes and antibodies; R. Ramirez-Solis for training regarding embryonic stem cells; C. Krull, D. Crotty and S. Fraser for discussions and help with hindbrain explant culture; E. Gngorieva for histology; C. Marcos and I. McKay for sharing unpublished data; A. Hewett for animal husbandry and J. Sharpe and H. Marshall for transgenic lines. This work was supported by the MRC, NATO Travel and EEC HCM network grants (R.K.) an MRC programme grant (A.L.) and an EEC HCM postdoctoral fellowship (M.S.).

CORRESPONDENCE and requests for material should be addressed to R.K. (e-mail: r.krumla@nimr.mrc.ac.uk).

CD22 regulates thymus-independent responses and the lifespan of B cells

Kevin L. Otipoby*, Kristin B. Andersson†||, Kevin E. Draves†, Stephen J. Klaus*, Andrew G. Farr*‡, James D. Kerner*, Roger M. Perlmutter*§, Che-Leung Law† & Edward A. Clark*‡

Departments of * Immunology, † Microbiology, ‡ Biological Structure, § Biochemistry and Medicine, and the Howard Hughes Medical Institute, University of Washington Medical Center, Box 357660, Seattle, Washington 98195, USA

THE B-lymphocyte-restricted glycoprotein CD22 is expressed on mature IgM⁺ IgD⁺ B cells^{1–3}, and is capable of binding to ligands on T and B cells^{2,4–7}. CD22 can interact with both the B-cell antigen receptor (BCR) complex^{8,9} and signalling molecules, including the protein tyrosine phosphatase SHP1 (PTPIC, SHP)^{10–12}, a putative negative regulator of BCR signalling^{13,14}. Thus CD22 may facilitate interactions with lymphocytes and regulate the threshold of BCR signalling^{1,11,13}. To define the *in vivo* function of CD22, we generated CD22-deficient mice. Here we show that CD22 is required for normal antibody responses to thymus-independent antigens and regulates the lifespan of mature B cells.

Mice homozygous for a deletion of the *Cd22* locus¹⁵ were produced by targeted disruption of the *Cd22^b* allele (Fig. 1a, b). Nearly all B220⁺ (CD45R⁺) splenocytes in wild-type mice expressed CD22, whereas B cells from CD22-mutant mice, as expected, did not (Fig. 1c). Overall, CD22-deficient mice were both grossly normal and fertile. Early B-cell subpopulations¹⁶, fractions A–D and fraction E (B220⁺ IgM⁺ IgD⁺) were present at normal numbers in the bone marrow of CD22-null mice (data not shown). However, mature B220⁺ IgM⁺ IgD⁺, long-lived recirculating B cells^{17,18}, were greatly reduced in CD22-deficient bone marrow (Fig. 1d). Mutant mice contained relatively normal

levels of peripheral T cells (CD3⁺, data not shown), B cells (B220⁺; Fig. 1c) and CD5⁺ cells (data not shown), which normally express CD22¹⁹. Splenic B cells from CD22-deficient mice consistently displayed decreased membrane IgM (mIgM) and increased expression of major histocompatibility complex (MHC) class II (Fig. 1e). This was not due to a general dysregulation of surface markers on CD22-negative B cells, which had, for example, normal levels of B220 (Fig. 1e). Generally, mIgD levels were identical on *Cd22*^{−/−} and *Cd22*^{+/+} mice (data not shown).

After ligating the mIgM BCR, CD22 is phosphorylated on tyrosine^{8,9,20}, potentially recruiting the tyrosine kinase Syk, phospholipase C-γ1 (PLCγ1), and SHP1 to CD22 (refs 10–12). Because B cells from SHP1-deficient mice display elevated calcium responses to antigen¹⁴, we analysed the ability of wild-type and mutant B cells to increase intracellular free calcium ([Ca²⁺]_i) after mIgM or mIgD crosslinking (Fig. 2a). Mutant B cells were about tenfold more sensitive to BCR engagement by graded doses of either anti-IgM or anti-IgD, even though they expressed less mIgM.

To test whether the increase in [Ca²⁺]_i on BCR engagement altered distal cellular responses in CD22-deficient B cells, we compared their *in vitro* proliferative response to B cells from wild-type mice (Fig. 2b). CD22-deficient B cells had decreased proliferative responses after BCR crosslinking compared with wild-type B cells; this decrease was not restored by adding exogenous interleukin (IL)-4. In contrast, no differences were evident after CD40 ligation in the absence (Fig. 2b) or presence (data not shown) of IL-4. Surprisingly, CD22-deficient B cells consistently showed a dramatically increased proliferative response to lipopolysaccharide; the basis for this is not yet known.

The normal proliferative response of mutant B cells *in vitro* to CD40 ligation (Fig. 2b) was mirrored in mutant mice by normal primary and secondary IgM and IgG responses to a thymus-dependent (TD) antigen, DNP-KLH (Fig. 3). The reduced proliferative response of CD22-deficient B cells was paralleled by significantly less antigen-specific IgM, IgG1 and IgG3 after challenge with the TI-2 antigen, DNP-Ficoll. The differences were greater for IgG1 and IgG3 responses than for IgM, suggesting that CD22 contributes to isotype switching in response to TI-2 antigens. Taken together these data suggest that CD22 expression is not required for T-cell-dependent B-cell growth and differentiation, but is required for normal antibody response to TI

|| Present address: Department of Biochemistry, University of Oslo, PB 1041 Blindern, N-0316, Oslo, Norway.

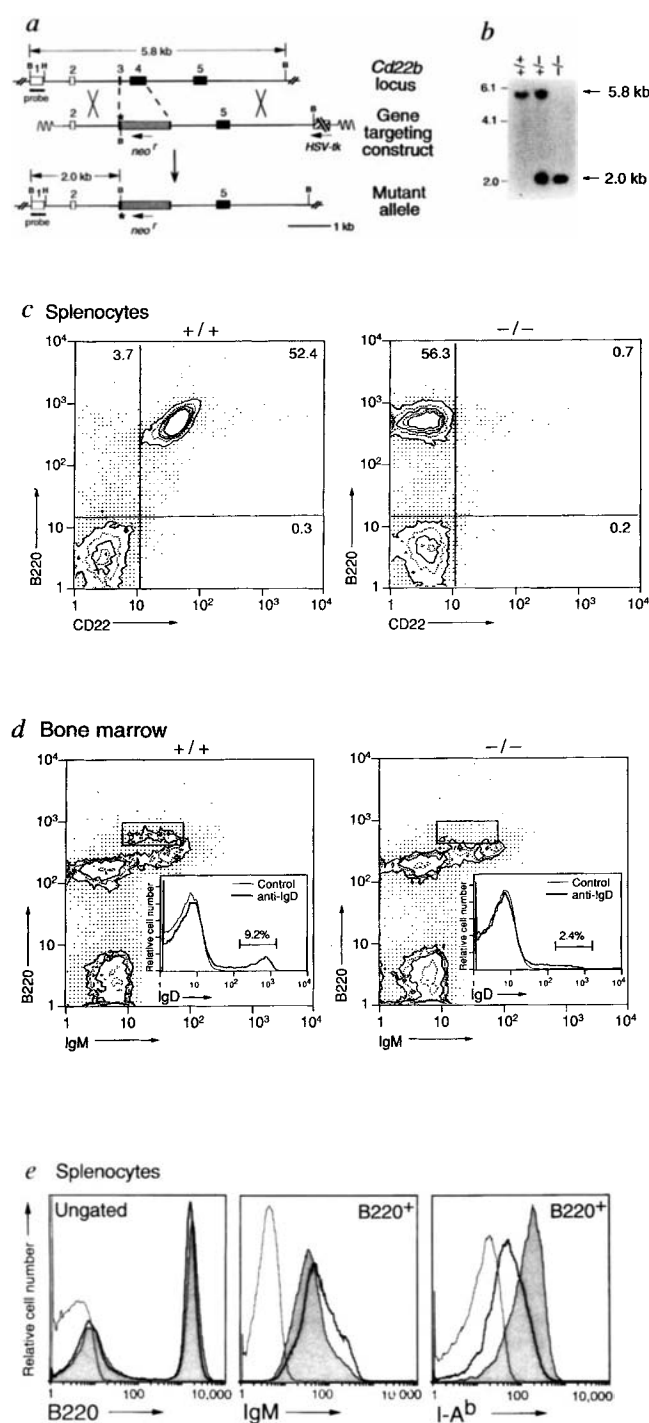


FIG. 1 Generation of *Cd22*^{-/-} mice. **a**, Configuration of wild-type *Cd22b* locus (top), gene-targeting construct (middle), and predicted mutant allele (bottom). Exons 1 and 2 (open boxes) encode 5' untranslated sequence, exon 3, the start codon and leader sequence, and exons 4 and 5, domains 1 and 2 of CD22. The probe used for Southern analysis and the lengths of the diagnostic restriction fragments are shown. B, *Bam*HI; H, *Hind*III. **b**, Wild-type, heterozygous and mutant animals displayed the predicted restriction patterns. **c**, Splenocytes from *Cd22*^{+/+} or *Cd22*^{-/-} mice were stained for CD22 and B220 expression. **d**, Bone-marrow cells obtained from the femurs of *Cd22*^{+/+} and *Cd22*^{-/-} mice were stained for B220, mIgM and mIgD. The B220^{hi} mIgM⁺ cells (boxed) and the mIgD⁺ population of mature, long-lived B cells (inset) were reduced. Three-colour analysis confirmed these were the same population. **e**, Splenocytes from *Cd22*^{+/+} (bold line) and *Cd22*^{-/-} (shaded area) mice were analysed for B220, mIgM and MHC class II expression. B220⁺ lymphocytes were used for mIgM and MHC class II overlays. Background staining of cells by isotype-matched control is shown (thin line).

antigens. The defective antibody production to a TI-2 antigen in CD22-deficient mice was not due to noticeable alterations in splenic architecture; a series of frozen spleen sections from wild-type and CD22-deficient mice either unimmunized or immunized with TNP-Ficoll (10 µg) or DNP-KLH (100 µg) revealed no differences after staining for CD3 or B220 expression or PNA binding (data not shown).

To examine why CD22-null B cells have defective proliferative and antibody responses, we compared the ability of wild-type and CD22-deficient B cells to enter the cell cycle or undergo apoptosis after anti-IgM stimulation (Fig. 4a). Both unstimulated wild-type and mutant B cells underwent significant apoptosis, which was slightly yet reproducibly elevated in CD22-deficient B cells. Although anti-IgM stimulation induced both wild-type and mutant B cells to enter the cell cycle, it triggered fewer CD22-deficient B cells to do so. This combination of enhanced apoptosis and decreased cell-cycle entry after BCR signalling suggested that CD22-deficient B cells may have a shorter lifespan than CD22-positive B cells *in vivo*. To test this possibility, CD22-deficient and CD22-positive mice were continuously fed bromodeoxyuridine (BrdU), and the proportion of labelled B cells in the bone marrow and spleen was measured (Fig. 4b, c). As previously shown¹⁷, only the long-lived IgD⁺ bone-marrow cells remained BrdU-deficient after 7 (data not shown) or 14 (Fig. 4b) days in wild-type mice; consistent with the lack of B220^{hi} cells in the bone marrow, this population was virtually absent in CD22-deficient mice. Furthermore, the proportion of long-lived B cells was dramatically reduced in the spleens of CD22-deficient compared with CD22-positive mice (Fig. 4c), as shown by a larger percentage of BrdU-positive cells at all time points during continuous feeding of BrdU (pulse). The proportion of BrdU-positive, short-lived B cells rapidly declined in both mutant and wild-type mice by week 4, one week after removal of BrdU (chase). The proportion of BrdU-positive B cells levelled off in wild-type mice between weeks 4 and 6 but continued to decline in CD22-deficient mice, suggesting that the amount of long-lived B cells are reduced.

B cells from CD22-deficient mice display many features similar to anergic B cells from double-transgenic mice expressing both a BCR specific for hen egg lysozyme (anti-HEL) and soluble HEL (sHEL). Both anti-HEL/sHEL and CD22-deficient mice have decreased expression of mIgM, defective antibody and proliferative responses after BCR stimulation¹³, and reduced numbers of long-lived B cells^{21,22}. B cells from anti-HEL/sHEL and *Cd22*^{-/-} mice differ in their Ca²⁺ response after BCR ligation, which are usually diminished in anergic B cells²³. This difference is probably caused by a lack of SHP1 recruitment to the antigen receptor complex in CD22-deficient B cells, because SHP1-deficient B cells also have increased BCR-induced Ca²⁺ mobilization¹⁴. A lack of SHP1 recruitment to the BCR complex in CD22-deficient B cells may result in low-level chronic BCR stimulation, so the threshold for anergy induction is reached at a lower degree of receptor ligation¹⁵. Peripheral B cells that develop in the absence of CD22 may downmodulate events distal to the BCR, resulting in a state of decreased responsiveness.

Although the molecular basis of the anergic-like phenotype in CD22-deficient mice is not clear, our data suggest that CD22 might function in the peripheral lymphoid tissues where newly formed B cells emigrating from the bone marrow die within a few days unless rescued by appropriate signals, presumably through both the BCR complex and molecules involved in cell-cell interactions^{13,22,24-26}. Thus, after newly formed B cells migrate to the outer T-cell zones near splenic and lymph-node follicles, CD22 might regulate the threshold of antigen-receptor signalling necessary for peripheral B-cell survival and recruitment into a long-lived pool within this competitive environment^{24,26}. The lack of CD22 may also affect the ability of B cells to locate in, or move to, niches that support long-lived cells.

A recent study²⁷ found, as we have, that CD22-mutant B cells have augmented Ca²⁺ release after BCR ligation, but these CD22-mutant mice had several other phenotypic differences compared

to our CD22-deficient mice. These differences may stem from the gene targeting constructs used; the CD22-mutant mice produced by O'Keefe *et al.*²⁷, unlike our CD22-deficient mice, may be producing secreted CD22 after splicing out of exons disrupted by *Neo*²⁸. □

Methods

Targeted disruption of the *Cd22* gene. A gene targeting construct was made from a 5.2-kb *Bam*HI fragment cloned from a 129/Ola genomic library by replacing the exons encoding the translational start site, leader sequence, ligand binding immunoglobulin domain 1, and the intervening intronic sequences¹⁵ with a neomycin resistance gene cassette²⁹ (*neo*). The splice junctions 5' of exon 3 and 3' of exon 4 were left intact and stop codons were introduced in all three reading frames immediately after the 5' border of exon 3 (Fig. 1a, asterisk). The *HSV-tk* gene³⁰ was added to allow for selection against random integration of the linearized construct. G418 and gancyclovir double-resistant clones of embryonic stem 14.1a cells transfected with the linearized

construct ($30 \mu\text{g ml}^{-1}$) were injected into 3.5-day blastocysts (C57BL/6). One chimaeric male was generated which was backcrossed 1–3 times to C57BL/6 females. Tail DNA was digested with *Bam*HI endonuclease and restriction fragments were detected by Southern analysis using the indicated *Bam*HI/*Hind*III fragment.

Flow cytometric analysis. Splenocytes were analysed by staining with anti-B220-PE (RA3-6B2; PharMingen), anti-CD22-biotin (2D6)¹⁵, anti-IgM-FITC (Bet-2), and anti-I-A^b-biotin (AF6-120.1; PharMingen). Streptavidin-FITC and streptavidin-PerCP were used as second-step reagents. Bone-marrow cells were stained with anti-B220-PE, anti-IgM-FITC, and anti-IgD-biotin (JA12.5), as described¹⁹. To monitor changes in $[\text{Ca}^{2+}]_i$, splenocytes (10^7 per ml) were loaded with Indo-1/AM (Calbiochem) dye ($7 \mu\text{g ml}^{-1}$) for 45 min at 37°C , stained with anti-CD3-FITC (2C11), and data collected for 45 s to establish baseline violet/blue ratios. Data were collected on gated $\text{CD}3^+$ lymphocytes for a total of 5 min and analysed for violet/blue ratios.

Cell proliferation assays. Splenic B cells from littermates 10–12 weeks of age were depleted of T cells with an antibody cocktail of Thy1.2 (J1j.10), CD4 (GK1.5), and CD8 (3.155) rat monoclonal antibodies, followed by anti-rat kappa (Mar 18.5) and agarose absorbed guinea-pig complement (Accurate

FIG. 2 Calcium responses and proliferation of *Cd22*^{-/-} B cells. **a**, Splenic B cells from *Cd22*^{+/+} and *Cd22*^{-/-} littermates were analysed for their ability to increase $[\text{Ca}^{2+}]_i$ on stimulation (arrows) with graded doses (in $\mu\text{g ml}^{-1}$) of goat anti-mouse IgM or goat anti-mouse IgD (Nordic). **b**, Purified B cells ($>90\%$ B220⁺, $<1\%$ CD3⁺ cells) from *Cd22*^{+/+} (solid line) and *Cd22*^{-/-} (broken line) littermates were incubated with anti-IgM, with or without IL-4, anti-CD40 or lipopolysaccharide (LPS) for 48 h. Similar results for the LPS response were seen in six different experiments using littermates from different breeding pairs. Values from quadruplicate cultures were used to calculate standard errors of the mean (s.e.m.).

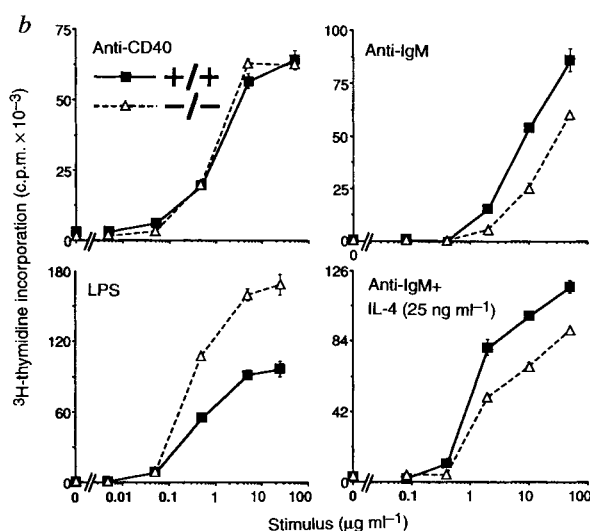
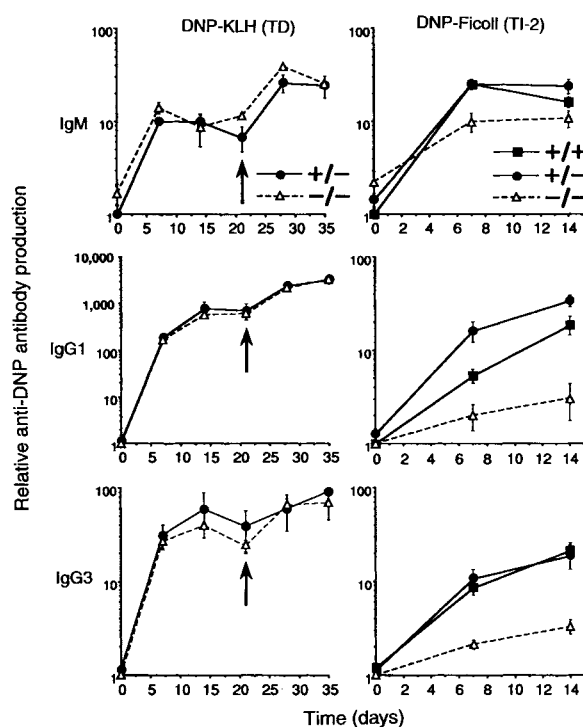
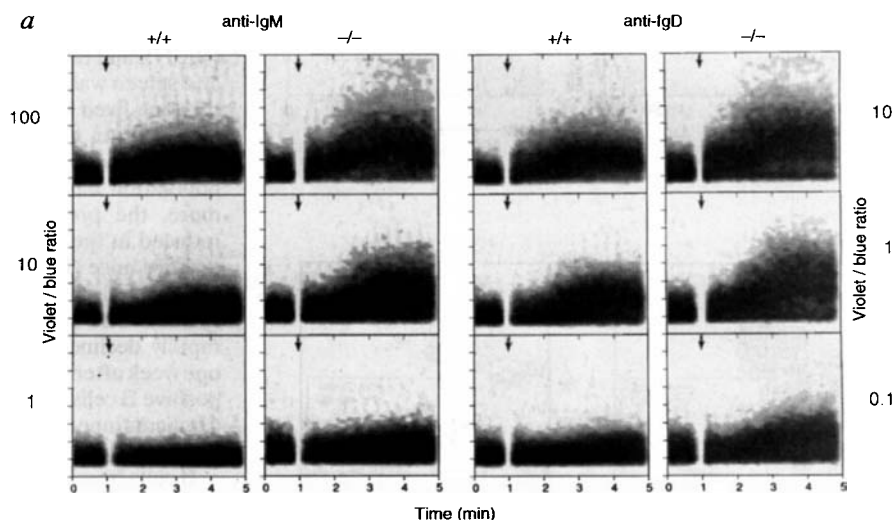


FIG. 3 Responses to TD and TI-2 antigens in *Cd22*^{-/-} mice. Means $\pm$ s.e.m., from 3 or 4 mice per genotype are shown for DNP-KLH (TD) and DNP-Ficoll (TI-2) immunizations, respectively. Arrows indicate the time of secondary immunization with DNP-KLH. In one experiment (2 mice per genotype), the antibody response to DNP-Ficoll was shown to be declining by 21 days postimmunization.

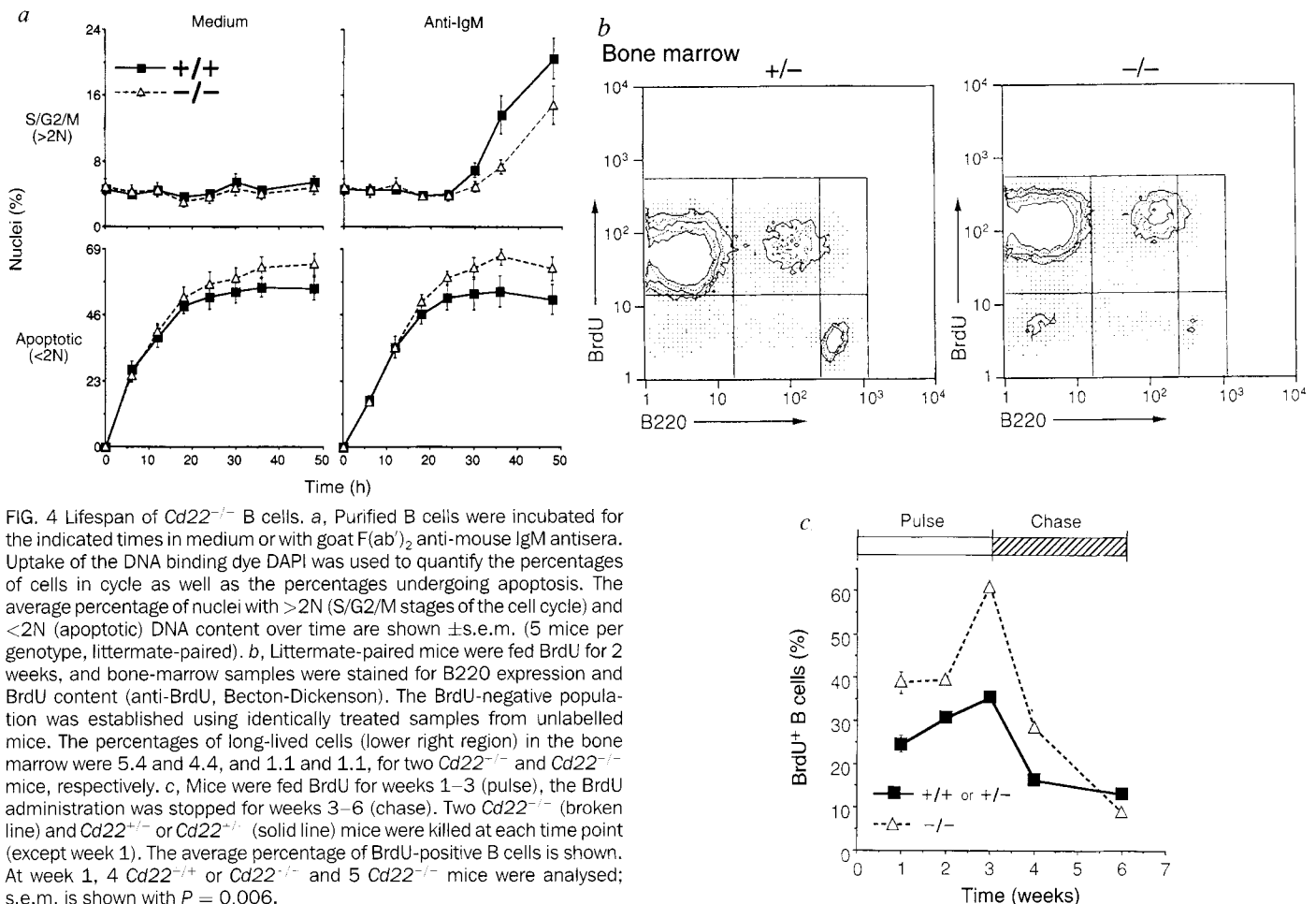


FIG. 4 Lifespan of $Cd22^{-/-}$ B cells. **a**, Purified B cells were incubated for the indicated times in medium or with goat $F(ab')_2$ anti-mouse IgM antisera. Uptake of the DNA binding dye DAPI was used to quantify the percentages of cells in cycle as well as the percentages undergoing apoptosis. The average percentage of nuclei with >2N (S/G2/M stages of the cell cycle) and <2N (apoptotic) DNA content over time are shown $\pm$ s.e.m. (5 mice per genotype, littermate-paired). **b**, Littermate-paired mice were fed BrdU for 2 weeks, and bone-marrow samples were stained for B220 expression and BrdU content (anti-BrdU, Becton-Dickenson). The BrdU-negative population was established using identically treated samples from unlabelled mice. The percentages of long-lived cells (lower right region) in the bone marrow were 5.4 and 4.4, and 1.1 and 1.1, for two $Cd22^{-/-}$ and $Cd22^{-/-}$ mice, respectively. **c**, Mice were fed BrdU for weeks 1–3 (pulse), the BrdU administration was stopped for weeks 3–6 (chase). Two $Cd22^{-/-}$ (broken line) and $Cd22^{+/+}$ or $Cd22^{+/-}$ (solid line) mice were killed at each time point (except week 1). The average percentage of BrdU-positive B cells is shown. At week 1, 4 $Cd22^{+/+}$ or $Cd22^{+/-}$ and 5 $Cd22^{-/-}$ mice were analysed; s.e.m. is shown with $P = 0.006$.

Chemical and Scientific). B cells were separated from buoyant macrophages, dead cells and debris by discontinuous Percoll gradient centrifugation and collected from the 53%/72% interface. Equivalent numbers of B cells of each phenotype cultured at 1.5×10^6 cells ml^{-1} in complete RPMI medium containing β -mercaptoethanol (5×10^{-5} M) and with the indicated concentrations of anti-CD40 (IC10), goat $F(ab')_2$ anti-mouse IgM antisera (Jackson), with or without 25 ng ml^{-1} IL-4 (R & D Systems), or lipopolysaccharide (Sigma). Proliferation was measured by 3H -thymidine incorporation (1 μ Ci per well) during the last 8 h of a 48-h incubation.

Immunizations and serum antibody assays. Littermate-paired mice were pre-bled and immunized intraperitoneally with either alum-precipitated DNP-KLH (100 μ g) or DNP-Ficol (10 μ g) in PBS. Serum was collected, and the relative anti-DNP titres were determined by ELISA. Briefly, ELISA plates were coated with DNP-BSA (4 μ g) and blocked with BSA (1%) and Tween-20 (0.05%). Sera were serially diluted in triplicate and detected with sheep anti-mouse isotype-specific antisera (450 ng ml^{-1}) (Binding Site) and donkey anti-sheep-horseradish peroxidase antisera (375 ng ml^{-1}) (Jackson). To control for plate-to-plate variation, anti-DNP antiserum was diluted in bulk and used as a relative standard on each plate. Values of the test samples at half-maximal optical density (OD) were normalized against the value at half-maximal OD of the standard dilutions on the corresponding plate.

Cell cycle and BrdU assays. Purified B cells were cultured with or without goat $F(ab')_2$ anti-mouse IgM antisera (5 μ g ml^{-1}). Cells were washed, resuspended in a $1 \times$ DAPI (4,6-diamidino-2-phenylindole) solution (2 mM $CaCl_2$, 22 mM $MgCl_2$, 50 μ g ml^{-1} BSA, 0.1% NP-40, 10 μ g ml^{-1} DAPI, and 10% DMSO), and frozen at $-70^\circ C$ until the time of analysis. BrdU labelling of cells *in vivo* and staining were performed³¹.

- Stamenkovic, I., Sgroi, D., Aruffo, A., Sy, M. S. & Anderson, T. *Cell* **66**, 1133–1144 (1991).
- Hanasaki, K., Varki, A., Stamenkovic, I. & Bevilacqua, M. P. *J. Biol. Chem.* **269**, 10637–10643 (1994).
- Law, C. L., Aruffo, A., Chandran, K. A., Doty, R. T. & Clark, E. A. *J. Immunol.* **155**, 3368–3376 (1995).
- Leprince, C., Draves, K. E., Geahlen, R. L., Ledbetter, J. A. & Clark, E. A. *Proc. Natl Acad. Sci. USA* **90**, 3236–3240 (1993).
- Peaker, C. J. & Neuberger, M. S. *Eur. J. Immunol.* **23**, 1358–1363 (1993).
- Campbell, M. A. & Klinman, N. R. *Eur. J. Immunol.* **25**, 1573–1579 (1995).
- Doody, G. M. *et al. Science* **269**, 242–244 (1995).
- Law, C. L. *et al. J. Exp. Med.* **183**, 547–560 (1996).
- Goodnow, C. C. *Proc. Natl Acad. Sci. USA* **93**, 2264–2271 (1996).
- Cyster, J. G. & Goodnow, C. C. *Immunity* **2**, 13–24 (1995).
- Law, C. L. *et al. J. Immunol.* **151**, 175–187 (1993).
- Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. *J. Exp. Med.* **173**, 1213–1225 (1991).
- Forster, I., Vieira, P. & Rajewsky, K. *Int. Immunol.* **1**, 321–331 (1989).
- Allman, D. M., Ferguson, S. E., Lentz, V. M. & Cancro, M. P. *J. Immunol.* **151**, 4431–4444 (1993).
- Andersson, K. B. *et al. Eur. J. Immunol.* (in the press).
- Schulte, R. J., Campbell, M. A., Fischer, W. H. & Sefton, B. M. *Science* **258**, 1001–1004 (1992).
- Fulcher, D. A. & Basten, A. J. *Exp. Med.* **179**, 125–134 (1994).
- Cyster, J. G. & Goodnow, C. C. *Immunity* **3**, 691–701 (1995).
- Cooke, M. P. *et al. J. Exp. Med.* **179**, 425–438 (1994).
- Cyster, J. G., Hartley, S. B. & Goodnow, C. C. *Nature* **371**, 389–395 (1994).
- Ramthell, J. C. *et al. Cell* **87**, 319–329 (1996).
- Fulcher, D. A. *et al. J. Exp. Med.* **183**, 2313–2328 (1996).
- O'Keefe, T. L., Williams, G. T., Davies, S. L. & Neuberger, M. S. *Science* **274**, 798–801 (1996).
- Muller, U. *et al. Cell* **79**, 755–765 (1994).
- Thomas, K. R. & Capecci, M. R. *Cell* **51**, 503–512 (1987).
- Mansour, S. L., Thomas, K. R. & Capecci, M. R. *Nature* **336**, 348–352 (1988).
- Tough, D. & Sprent, J. *J. Exp. Med.* **179**, 1127–1135 (1994).

ACKNOWLEDGEMENTS. K.L.O. and K.B.A. contributed equally to this work. We thank R. Iderda and R. Torres for assistance in setting up the gene disruption; K. Forbush for help with animals; and R. Palmiter for the 129/Ola library. R.M.P. is an investigator of the Howard Hughes Medical Institute. This work was supported by grants from the NIH (E.A.C.) and by a Howard Hughes Medical Institute pre-doctoral fellowship (K.L.O.).

CORRESPONDENCE and requests for materials should be addressed to E.A.C. (e-mail: eclark@bart.prc.washington.edu.).

Received 23 September; accepted 26 November 1996.

- Law, C. L., Sidorenko, S. P. & Clark, E. A. *Immunol. Today* **15**, 442–449 (1994).
- Stamenkovic, I. & Seed, B. *Nature* **345**, 74–77 (1990).
- Wilson, G. L., Fox, C. H., Fauci, A. S. & Kehrl, J. H. *J. Exp. Med.* **173**, 137–146 (1991).
- Torres, R. M. *et al. J. Immunol.* **149**, 2641–2649 (1992).

NMR structure and mutagenesis of the Fas (APO-1/CD95) death domain

Baohua Huang, Matthias Eberstadt, Edward T. Olejniczak, Robert P. Meadows & Stephen W. Fesik

Pharmaceutical Discovery Division, Abbott Laboratories, Abbott Park, Illinois 60064, USA

PROGRAMMED cell death (apoptosis) mediated by the cytokine receptor Fas is critical for the removal of autoreactive T cells¹, the mechanism of immune privilege^{2,3}, and for maintenance of immune-system homeostasis⁴. Signalling of programmed cell death involves the self-association of a conserved cytoplasmic region of Fas called the death domain^{5–7} and interaction with another death-domain-containing protein, FADD⁸ (also known as MORT1)⁹. Although death domains are found in several proteins¹⁰, their three-dimensional structure and the manner in which they interact is unknown. Here we describe the solution structure of the Fas death domain, as determined by NMR spectroscopy. The structure consists of six antiparallel, amphipathic α -helices arranged in a novel fold. From the structure and from site-directed mutagenesis, we have identified the region of the death domain involved in self-association and binding to the downstream signalling partner FADD.

Structural studies of death domains have been hampered by the propensity of this class of proteins to self-associate and form large aggregates in solution. At pH 7.0, the Fas death domain aggre-

gates at the concentrations necessary for NMR structure determination. However, the protein is predominantly monomeric at concentrations as high as 1.0 mM at pH 4.0, as shown by light scattering (data not shown). Under these conditions, circular dichroism (CD) spectra indicate that the Fas death domain maintains a fold similar to that found in the protein at higher pH. We also determined the secondary structure of a non-aggregating mutant protein (D244A) by NMR at pH 6.0 and found it to be the same as the wild-type protein at pH 4.0 (data not shown).

The structure of a portion of Fas encompassing the death domain (residues 202–319; Fig. 1a) was determined from a total of 1,328 NMR-derived restraints. The backbone and hydrophobic side chains within the core of the protein were well defined by the NMR data for residues 210–305 (Fig. 1b and Table 1). Interestingly, this region of the protein corresponds to the boundaries of the death domain determined by sequence homology¹⁰ and deletion mutagenesis of the tumour-necrosis factor receptor (TNF-R1)⁵ (Fig. 1a). The fifteen carboxy-terminal residues of Fas (305–319), which play a negative regulatory role⁶ and interact with the protein tyrosine phosphatase FAP-1 (ref. 11), are disordered.

The structure of the Fas death domain consists of six amphipathic α -helices arranged antiparallel to one another (Fig. 1c, d), with the side chains of the hydrophobic residues forming an extensive network of interactions that constitute the core of the protein (Fig. 1b). Helices 1 and 2 are centrally located, with α 3 and α 4 on one side and α 5 and α 6 on the other (Fig. 1d). An unusual crossing pattern is seen in which the loop connecting helices 4 and 5 bisects helices 1 and 2 (Fig. 1d). Using the program Deja Vu¹², no proteins of similar structure (<5.0 Å) were found when strict conservation of helical directionality and molecular connectivity were used in the search.

Only one notable hydrophobic site is located on the surface of

TABLE 1 Structural statistics and r.m.s.ds for the NMR-derived structures of the Fas death domain*

Structural statistics	(SA)	(SA) _r
R.m.s. deviation from experimental distance restraints (Å)†		
Intra-residue (377)	0.006 ± 0.002	0.006
Sequential (354)	0.021 ± 0.002	0.014
Medium-range (208)	0.012 ± 0.001	0.013
Long-range (273)	0.016 ± 0.001	0.014
Hydrogen bonds (58)	0.017 ± 0.002	0.013
R.m.s. deviation from experimental torsional angle restraints (degrees)‡		
Φ angles (58)	0.05 ± 0.04	0.07
X-PLOR potential energies (kcal mol ⁻¹)		
E _{tot}	59 ± 2	68
E _{bond}	4 ± 0.2	5
E _{ang}	25 ± 1	31
E _{imp}	6 ± 0.1	5
E _{repel}	14 ± 1	18
E _{noe}	9 ± 1	9
E _{cdih}	0.01 ± 0.1	0.01
E _{LJ} §	-612 ± 11	-578
Cartesian coordinate r.m.s.d. (Å)	N, C [*] , and C'	All heavy
(SA) versus (SA) _r	0.55 ± 0.10	1.1 ± 0.11

* (SA) is the ensemble of 20 NMR-derived solution structures encompassing the Fas death domain; (SA)_r is the mean atomic structure obtained by averaging the individual (SA) structures following a least-squares superposition of the backbone heavy atoms for residues 210–305; (SA)_r is the energy-minimized average structure. The X-PLOR F_{repel} function²⁶ was used to simulate van der Waals interactions with a force constant of 4.0 kcal mol⁻¹ Å⁻⁴, with atomic radii set to 0.8 times their CHARMM²⁷ values.

† Distance restraints were used with a square-well potential (F_{noe} = 50 kcal mol⁻¹ Å⁻²). Hydrogen bonds were included as distance restraints and given bounds of 1.8–2.3 Å (H–O) and 2.8–3.3 Å (N–O). Medium-range nuclear Overhauser effects (NOEs) are observed between protons separated by >1 and >5 residues in the primary sequence. Long-range NOEs are observed between protons separated by five or more residues. A total of 1,212 non-trivial NOE-derived distance restraints was used. No distance restraint was violated by more than 0.25 Å in any of the final structures.

‡ Torsional restraints were applied to Φ angles with values of -120 ± 40° for those angles with ³J_{HNH_α} coupling constants larger than 9.0 Hz, and -60 ± 40° for coupling constants <5.5 Hz. Restraints for the latter were applied only in α -helical regions. Force constants of 200 kcal mol⁻¹ rad⁻² were applied for all torsional restraints. No torsional angle restraint was violated by more than 5° in any of the final structures.

§ The Lennard-Jones potential was not used during the refinement stage.

|| R.m.s.d. for residues 210–305.

the protein (Fig. 2a). However, many charged residues are present, suggesting that electrostatic interactions may mediate molecular interactions involving the death domain. While optimizing the solution conditions for NMR, we found that self-aggregation of the Fas death domain is decreased at high salt concentrations (1 M NaCl) and at low pHs (<5.0), where acidic residues become protonated.

The death domains found in other proteins such as TNF-R1 (ref. 5), FADD⁸, TRADD¹³ and RIP¹⁴ are homologous to Fas¹⁰ and probably adopt similar three-dimensional structures. In con-

trast, Reaper, a 65-residue protein that mediates programmed cell death in *Drosophila*¹⁵, may not be structurally similar to the death domain, as was suggested¹⁶. Based on sequence alignment¹⁶, Reaper is predicted to contain a proline in the middle of $\alpha 5$ and lack $\alpha 1$, $\alpha 6$ and part of $\alpha 2$.

To determine which regions of the death domain are responsible for self-association and interaction with the downstream signalling protein FADD, site-directed mutant proteins were prepared and tested for their ability to aggregate and bind to FADD. Based on the structure of Fas, the hydrophobic site

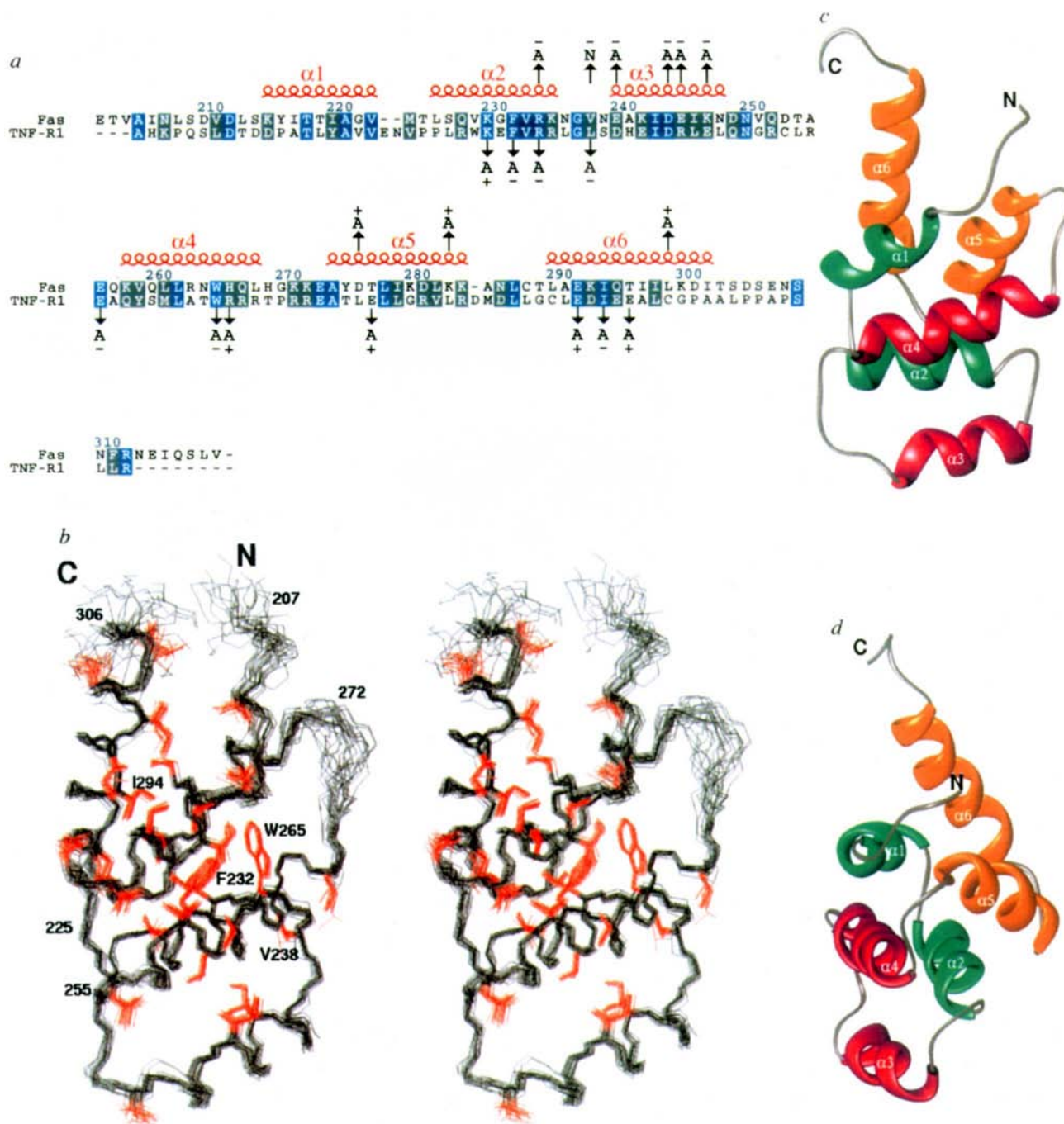


FIG. 1 a, Sequence alignment of the death domains of Fas and TNF-R1. The numbering in Fas is as described in ref. 30. Identical residues are highlighted in blue, and homologous residues in grey. The helices in the NMR structure of the Fas death domain are indicated. Mutations are indicated by an arrow pointing to the new amino acid (one-letter code). A minus sign immediately below the amino acid for TNF-R1 indicates a loss in

TNF-R1-mediated cytotoxicity⁹. A minus sign above the Fas mutation indicates that this mutant aggregates less than the wild-type protein. b, Stereoview of the backbone (N, C α , C') of 20 superimposed NMR-derived structures of the Fas death domain. The side chains of residues in the hydrophobic core are coloured red. c d, Ribbons²⁸ depictions of the averaged minimized NMR structure of the Fas death domain.

formed by $\alpha 5$ and $\alpha 6$ is a potential binding site (Fig. 2b). However, mutations of residues in this region (L299A and D276A) did not alter the mutant protein's ability to interact with itself (Fig. 3a) or FADD (Fig. 3b). In contrast, other mutations of the Fas death domain, including a naturally occurring mutation in Fas (V238N) that leads to a lymphoproliferation (*lpr*) syndrome in mice¹⁷, greatly reduced self-association and binding to FADD compared with the wild-type protein (Fig. 3). These residues are located in $\alpha 2$ (R234A), in the loop between $\alpha 2$ and $\alpha 3$ (V238N), and in $\alpha 3$ (E240A and D244A). Other mutant proteins (E245A and K247A) did not readily self-associate (Fig. 3a), but still interact with FADD (Fig. 3b), suggesting that aggregation of Fas is not a requirement for FADD binding.

Mutations of the Fas death domain may block protein-protein interactions directly by altering a residue that is critical for protein binding, or indirectly by disrupting the structure of the protein. To determine whether the mutations of the Fas death domain disrupt

the overall fold, we compared the $^{15}\text{N}/^1\text{H}$ amide chemical shifts of the mutants to the wild-type protein. For most of the mutant proteins, the amide chemical shifts for only a few residues near the site of the mutation were found to be different from the wild-type protein. Therefore mutations of these residues do not disrupt the protein structure but must directly affect protein/protein interactions involving the Fas death domain. In contrast, we observed many significant differences in amide chemical shifts for the *lpr* mutant (V238N) compared with the wild-type protein, suggesting that this mutation alters the protein structure. This is consistent with the location of this residue within the interior of the protein (Fig. 1b).

In previous studies based on alanine scanning mutagenesis⁵, several residues in the TNF-R1 death domain that are conserved in Fas were shown to be critical for TNF-R1 function (Fig. 1a). These residues are scattered throughout the protein, suggesting that interactions of the TNF-R1 death domain may involve many non-contiguous portions of the molecule. However, four of these

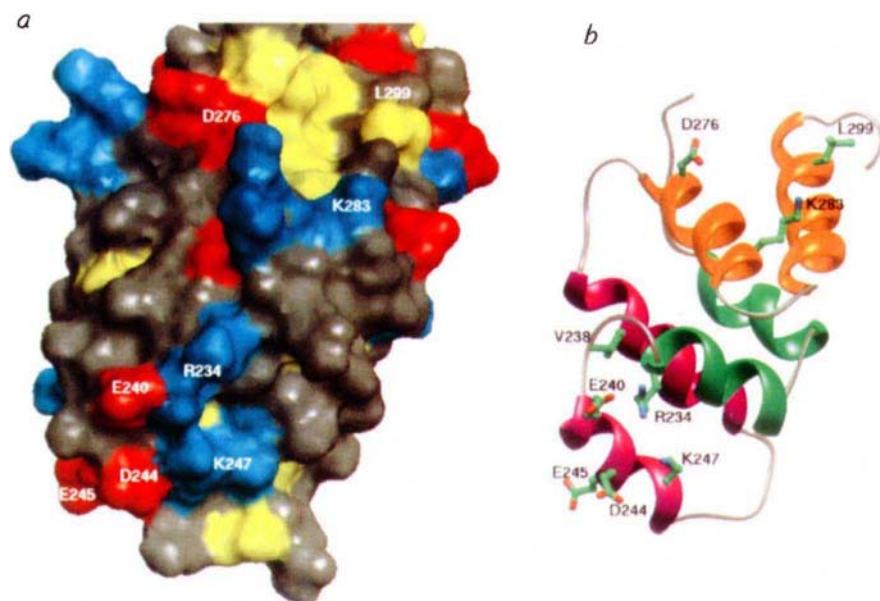


FIG. 2 a, Surface representation of the averaged minimized NMR structure of the Fas death domain. Lys and Arg side-chain atoms are shown in blue; Asp and Glu side-chain atoms are red; Phe, Val, Leu, Ile, Trp and Tyr side-chain atoms are yellow. b, Ribbons²⁸ representation of the averaged minimized NMR structure of the Fas death domain. Side-chain atoms of the mutated residues are colour-coded by atom type.

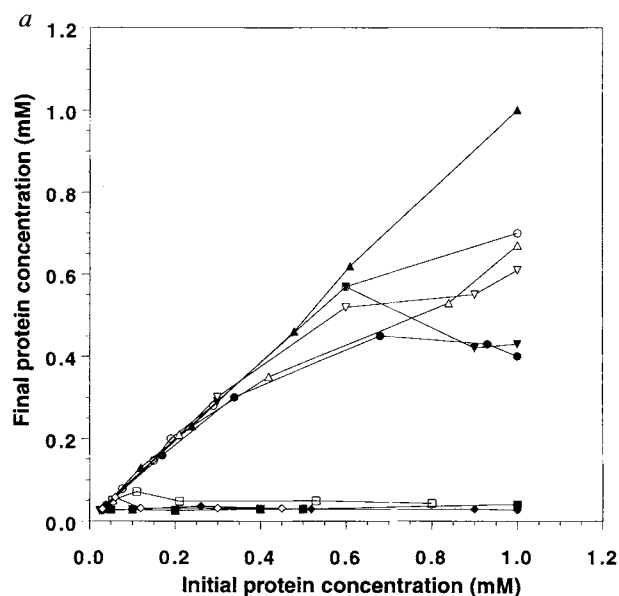
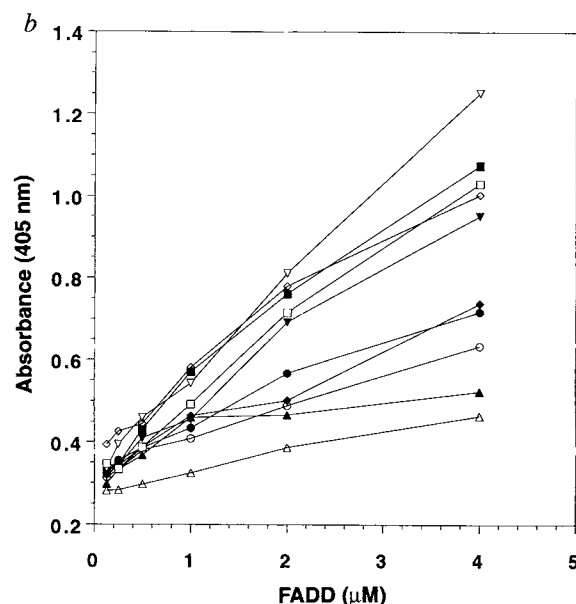


FIG. 3 a, Self-association of the Fas mutants. Mutant proteins were tested for their tendency to aggregate at different protein concentrations (see Methods). Symbols are the same in a and b. ■, Wild type; △, R234A; ○, V238N; ●, E240A; ▲, D244A; ▼, E245A; ▽, K247A; ◇, D276A; ◆,



K283A; □, L299A. b, Fas-FADD binding. Binding of Fas mutants to FADD was measured by an enzyme-linked assay similar to ELISA. The absorbance at 405 nm, representing the amount of bound FADD, is plotted against FADD concentration.

residues correspond to amino acids in Fas (F232, V238, W265 and I294) whose side chains are buried in the hydrophobic core (Fig. 1b) and which when mutated will probably alter the protein structure. Conversely, mutations of the charged residues in TNF-R1 that block function correspond to surface-exposed amino acids in Fas. These residues are located in the same region of the protein that we have shown to participate in interactions involving Fas. Thus, not only are the structures of this family of protein modules likely to be similar to the one shown here for the Fas death domain, but they may also mediate protein-protein interactions in a similar way. □

Methods

NMR sample preparation. The cDNA coding for the Fas cytoplasmic region was cloned from human fetal liver poly(A)⁺ RNA by RT-PCR. The Fas death domain (residues 202–319) was expressed in *Escherichia coli* using the pET30b vector (Novagen). The recombinant protein contained an extra methionine at the N terminus and a poly-histidine tag (–LEHHHHH) at the C terminus to facilitate purification. For the NMR experiments, uniformly ¹⁵N- or ¹³C-labelled protein was prepared by growing the *E. coli* strain HMS174(DE3) overexpressing the Fas death domain on M9 medium containing ¹⁵NH₄Cl with or without [U-¹³C]-glucose. A uniformly ¹⁵N, ¹³C-labelled and fractionally deuterated protein sample was also prepared by growing the cells in the presence of 75% ²H₂O. The recombinant Fas death domain was purified by affinity chromatography on a nickel-IDA column (Invitrogen). NMR samples contained 1 mM protein in 50 mM sodium acetate (pH 4.0) and 50 mM NaCl in H₂O/²H₂O (9/1) or ²H₂O.

Structure determination. All NMR spectra were acquired at 30 °C on a Bruker DMX500 or AMX600 NMR spectrometer. For assignment of the ¹H, ¹³C and ¹⁵N resonances of the backbone, a series of deuterium-decoupled triple-resonance experiments (HNCA, HN(CO)CA, HN(CA)CB and HN(COCA)CB)¹⁸ were recorded using the uniformly ¹⁵N, ¹³C-labelled and fractionally deuterated protein. From 3D HCCCH-TOCSY¹⁹ and HBHA(CO)NH¹⁹ experiments on the [U-¹⁵N, ¹³C]-labelled sample and from HC(CO)NH-TOCSY experiments on the [U-¹⁵N, ¹³C]-labelled and fractionally deuterated sample, ¹³C and ¹H resonances of the side chain were assigned. Stereospecific assignment of methyl groups of the Val and Leu residues were obtained using a fractionally ¹³C-labelled sample as described²⁰. Distance restraints were obtained from ¹⁵N- or ¹³C-resolved 3D NOESY experiments¹⁹; ϕ angle restraints were based on the ³J_{HN,H α} coupling constants measured in an HNHA experiment²¹. Slowly exchanging amide protons were identified from a series of ¹H/¹⁵N HSQC spectra recorded after the H₂O buffer was changed to a ²H₂O buffer. Structures of the Fas death domain were calculated with a distance geometry/simulated annealing protocol²² using the X-PLOR program²³. The structure calculations employed 1,212 proton/proton distance restraints, 58 hydrogen-bond distance restraints derived for 29 slowly exchanging amide protons, and 58 ϕ angle restraints. The NOE-derived restraints were categorized as strong (1.8–3 Å), medium (1.8–4 Å), or weak (1.8–5 Å) on the basis of the observed NOE intensities.

Site-directed mutagenesis. Mutant proteins were prepared using either the Chameleon double-stranded, site-directed mutagenesis kit or the Quick-change site-directed mutagenesis kit (Stratagene). The coding region was sequenced in each case.

Self-aggregation assay. Solutions of varying concentrations of wild-type and mutant Fas proteins were prepared in 50 mM potassium phosphate (pH 7.0), 0.3 M NaCl and 5 mM 2-mercaptoethanol. After incubating overnight at 25 °C, samples were microfuged for 2 h, and the protein concentration in the supernatant was determined by Bradford assay²⁴. The initial protein concentration was plotted against the final concentration as an indication of the tendency to aggregate.

Fas–FADD binding assay. Fas–FADD binding was determined by an enzyme-linked assay similar to an enzyme-linked immunosorbent assay (ELISA). Plates were coated overnight at 4 °C with Fas proteins (1 µg per well) diluted in carbonate buffer (0.1 M, pH 9.8). After washing four times with PBS containing 0.2% Tween-20, plates were incubated overnight at 4 °C with cell lysates containing S-tagged FADD in a series of dilutions. The bound S-tagged²⁴ FADD proteins were then measured with S-protein alkaline phosphatase (Novagen) at 4 °C. *p*-Nitrophenyl phosphate was used for colorimetric measurements, and absorbance was measured at 405 nm. FADD protein was prepared by synthesizing the gene coding for the amino-acid sequence from 10 oligonucleotides (each 90 nucleotides in length, with 30 nucleotides overlap among adjacent ones) as described²⁵. The synthetic FADD gene was cloned into pET29b (Novagen) and expressed in *E. coli* with an S-tag at the N terminus. The cleared cell lysate containing recombinant S-tagged FADD in PBS containing 5 mM 2-mercaptoethanol was used for the Fas–FADD binding assay.

- Griffith, T. S., Brunner, T., Fletcher, S. M., Green, D. R. & Ferguson, T. A. *Science* **270**, 1189–1192 (1995).
- Nagata, S. & Goldstein, P. *Science* **267**, 1449–1456 (1995).
- Tartaglia, L. A., Ayres, T. M., Wong, G. H. W. & Goeddel, D. V. *Cell* **74**, 845–853 (1993).
- Itoh, N. & Nagata, S. *J. Biol. Chem.* **268**, 10932–10937 (1993).
- Boldin, M. P. et al. *J. Biol. Chem.* **270**, 387–391 (1995).
- Chinnaiyan, A. M., O'Rourke, K., Tewari, M. & Dixit, V. M. *Cell* **81**, 505–512 (1995).
- Boldin, M. P. et al. *J. Biol. Chem.* **270**, 7795–7798 (1995).
- Feinstein, E., Kimchi, A., Wallach, D., Boldin, M. P. & Varfolomeev, E. *Trends Biochem. Sci.* **20**, 342–344 (1995).
- Sato, T., Irie, S., Kitada, S. & Reed, J. C. *Science* **268**, 411–415 (1995).
- Kleywegt, G. J. & Jones, T. A. *Structure* **3**, 535–540 (1995).
- Hsu, H., Xiong, J. & Goeddel, D. V. *Cell* **81**, 535–540 (1995).
- Stanger, B. Z., Leder, P., Lee, T. L., Kim, E. & Seed, B. *Cell* **81**, 513–523 (1995).
- White, K. et al. *Science* **264**, 677–683 (1994).
- Goldstein, P., Marguet, D. & Depraetere, V. *Cell* **84**, 185–186 (1995).
- Watanabe-Fukunaga, R., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Nagata, S. *Nature* **356**, 314–317 (1992).
- Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. *J. Am. Chem. Soc.* **116**, 11655–11666 (1994).
- Clare, G. M. & Gronenborn, A. M. *Meth. Enzymol.* **239**, 349–363 (1994).
- Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. *Biochemistry* **28**, 7510–7516 (1989).
- Vuister, G. W. & Bax, A. *J. Am. Chem. Soc.* **115**, 7772–7777 (1993).
- Kuszewski, J., Nilges, M. & Brünger, A. T. *J. Biomol. NMR* **2**, 33–56 (1992).
- Brünger, A. T. *X-PLOR 3.1 Manual* (Yale Univ. Press, New Haven, 1992).
- Kim, J.-S. & Raines, R. T. *Protein Sci.* **2**, 348–356 (1993).
- Dillon, P. J. & Rosen, C. A. *BioTechniques* **9**, 298–300 (1990).
- Nilges, M., Clare, G. M. & Gronenborn, A. M. *FEBS Lett.* **229**, 317–324 (1988).
- Brooks, B. R. et al. *J. Comput. Chem.* **4**, 187–217 (1983).
- Carson, M. J. *J. Mol. Graph.* **5**, 103–106 (1987).
- Bradford, M. M. *Anal. Biochem.* **72**, 248–254 (1976).
- Itoh, N. et al. *Cell* **66**, 233–243 (1991).

ACKNOWLEDGEMENTS. We thank E. Matayoshi for performing the light-scattering experiments, J. Harlan for recording the CD spectra, and J. E. Clampton and T. H. Lubben for assistance with the Fas–FADD binding assay.

CORRESPONDENCE and requests for materials should be addressed to S.W.F. (e-mail: fesik@steves.abbott.com). Coordinates for the NMR structure of the Fas death domain have been deposited in the Brookhaven Protein Data Bank under the accession code 1DDF.

The CBP co-activator is a histone acetyltransferase

Andrew J. Bannister & Tony Kouzarides

Wellcome/CRC Institute and Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK

THE CBP protein acts as a transcriptional adaptor for many different transcription factors by directly contacting DNA-bound activators^{1–7}. One mechanism by which CBP is thought to stimulate transcription is by recruiting the histone acetyltransferase (HAT) P/CAF to the promoter⁸. Here we show that CBP has intrinsic HAT activity. The HAT domain of CBP is adjacent to the binding site for the transcriptional activator E1A. Although E1A displaces P/CAF from CBP, it does not disrupt the CBP-associated HAT activity. Thus E1A carries HAT activity when complexed with CBP. Targeting CBP-associated HAT activity to specific promoters may therefore be a mechanism by which E1A acts as a transcriptional activator.

To monitor the HAT activity associated with CBP *in vivo*, we developed an immunoprecipitation–HAT (IP–HAT) assay. Immunoprecipitation of CBP from whole-cell extracts (and nuclear extracts; data not shown) brings down an associated acetylase activity that is specific for histones (Fig. 1). Antibodies directed against a variety of nuclear proteins only precipitate a background nonspecific activity, indicating that the CBP IP–HAT assay monitors acetylase activity complexed specifically with CBP.

One possibility is that CBP-bound HAT activity could be attributed to P/CAF, which forms a complex with CBP. If this were the case then E1A would be expected to diminish the CBP-associated HAT activity, because E1A displaces P/CAF from CBP⁸. However, transfection of COS cells with E1A or E1A mutants that lack the CBP binding site (E1AΔCR1 and CBP mut; ref. 5) does not affect CBP-bound HAT activity (Fig. 2a). To make sure that this was not a reflection of transfection efficiency,

Received 21 October; accepted 13 November 1996.

- Ju, S.-T. et al. *Nature* **373**, 444–448 (1995).
- Bellgrau, D. et al. *Nature* **377**, 630–632 (1995).

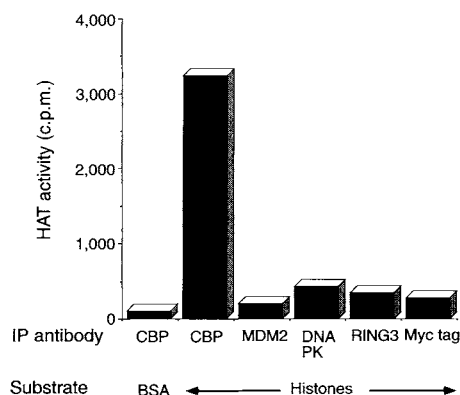


FIG. 1 CBP antibodies immunoprecipitate HAT activity in an IP-HAT assay. Immunoprecipitations were performed from Cos whole-cell extracts with the antibodies indicated. After extensive washing, the immune complexes were tested for their ability to acetylate either free histones or BSA.

we used a cell line that has an inducible E1A gene, which can be made to express high levels of E1A following CdCl_2 treatment⁹. Even under conditions where every cell is expressing E1A, the CBP-associated HAT activity is not diminished, and is actually slightly elevated (Fig. 2b).

These experiments suggested that the CBP may be associated with an additional HAT activity distinct from that of P/CAF. Because E1A does not displace this P/CAF-independent activity, we investigated whether E1A carries such activity after forming a complex with CBP. An E1A-specific antibody can precipitate low levels of HAT activity from extracts derived from the E1A-expressing MT-3 cells, but not from the parental GM637 cells, which do not contain E1A (Fig. 3a). Induction of E1A in MT-3 cells with CdCl_2 increases the amount of HAT activity associated with the E1A immunoprecipitate.

The HAT activity associated with the E1A immunoprecipitate is unlikely to be a property of E1A itself, because E1A does not display intrinsic HAT activity in liquid or gel activity assays (data not shown). We therefore investigated whether the E1A-associated HAT activity was a result of E1A binding CBP. For this analysis we used E1A mutants lacking the CBP binding site and tested their ability to associate with HAT activity. When an E1A-specific antibody is used in an IP-HAT assay, E1A-bound HAT activity is detected in extracts from cells transfected with E1A, but not in untransfected, control extracts (Fig. 3b). E1A mutants (ΔCR1 and CBP mut; ref. 5) that do not interact with p300/CBP *in vivo*¹⁰ are severely impaired in their ability to associate with HAT activity. In contrast, an E1A mutant that destroys the RB binding site in conserved region 1 (RB mut; ref. 5) still associates

FIG. 3 E1A binds HAT activity in a CBP-binding site-dependent manner. *a*, E1A antibodies were used to immunoprecipitate E1A complexes from control cells (GM637 cells) or an E1A-inducible cell line (MT-3 cells), either in the absence (–) or presence (+) of $5 \mu\text{M}$ CdCl_2 , which induces E1A expression. After extensive washing the immune complexes were tested for their ability to acetylate free histones. *b*, Cos cells were transiently transfected with $15 \mu\text{g}$ of the indicated RSV promoter-driven E1A expression plasmids. Whole-cell extracts were subjected to an E1A IP-HAT assay 48 h after transfection. All E1A proteins were expressed to similar levels as determined by western blotting. Top, a schematic representation of E1A indicating the position of the mutated RB and CBP binding sites.

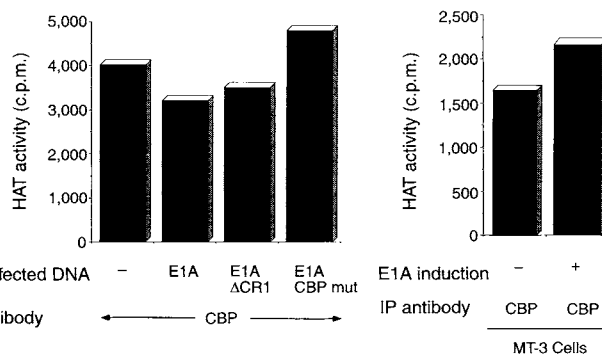
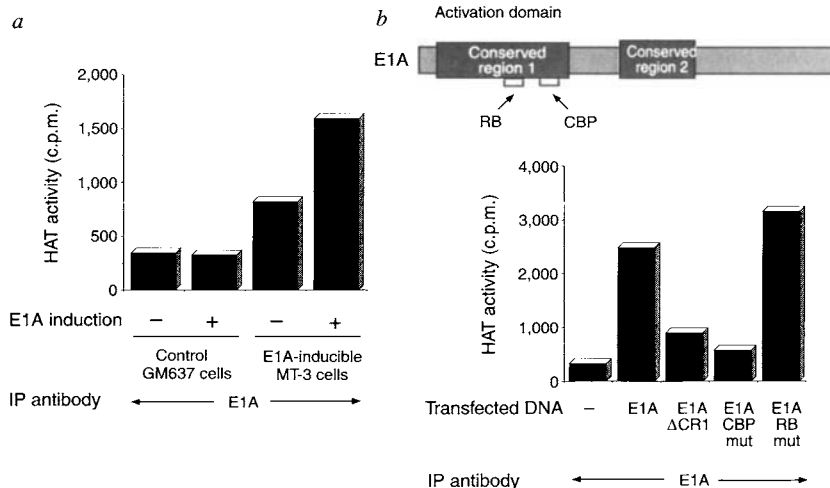


FIG. 2 E1A does not disrupt CBP-associated HAT activity. *a*, Cos cells were transiently transfected with $15 \mu\text{g}$ of the indicated RSV promoter-driven E1A expression plasmids. Whole-cell extracts were prepared 48 h after transfection, and a CBP IP-HAT assay was carried out. All the E1A proteins were expressed to similar levels, as determined by western blotting. *b*, CBP antibodies were used to immunoprecipitate CBP complexes from an E1A-inducible cell line (MT-3), either in the absence (–) or presence (+) of $5 \mu\text{M}$ CdCl_2 , which induces E1A expression⁹. After extensive washing the immune complexes were tested for their ability to acetylate free histones.

with the HAT activity.

These results indicate that the E1A-bound HAT activity is associated with CBP. We next tested whether CBP itself has acetyltransferase activity. A specific bacterially expressed domain of CBP (1,099–1,877) does have intrinsic HAT activity (Fig. 4). The CBP acetyltransferase activity is specific for histones, as BSA does not act as a substrate (data not shown). This domain contains two cysteine-rich motifs, ZZ and TAZ (ref. 11), which overlap the binding site for E1A and several cellular transcriptional activators. Removal of the TAZ motif does not affect the CBP HAT activity, but abolishes the binding of E1A to CBP (ref. 8 and data not shown). Thus E1A does not bind CBP residues that mediate HAT activity, which is consistent with the finding that the E1A–CBP complex retains HAT activity (Fig. 3b). Nucleosomes are a physiological target for HAT activity, so we tested the ability of CBP to acetylate histones within nucleosomes. All four nucleosomal histones are substrates for CBP HAT activity (Fig. 4c). This activity of CBP is specific for histones given that a panel of nuclear proteins (TFIIB, ORC2, CDC6, KU70/80, VP16) and nonspecific proteins (GST, BSA) are not substrates under the same SDS-PAGE (polyacrylamide gel electrophoresis) analysis (data not shown).

The HAT domain of CBP does not show any overt sequence similarity to the HAT domain of the GCN5 family of proteins to which P/CAF belongs. This raises the possibility that CBP can accomplish distinct functions by using two different HAT activ-

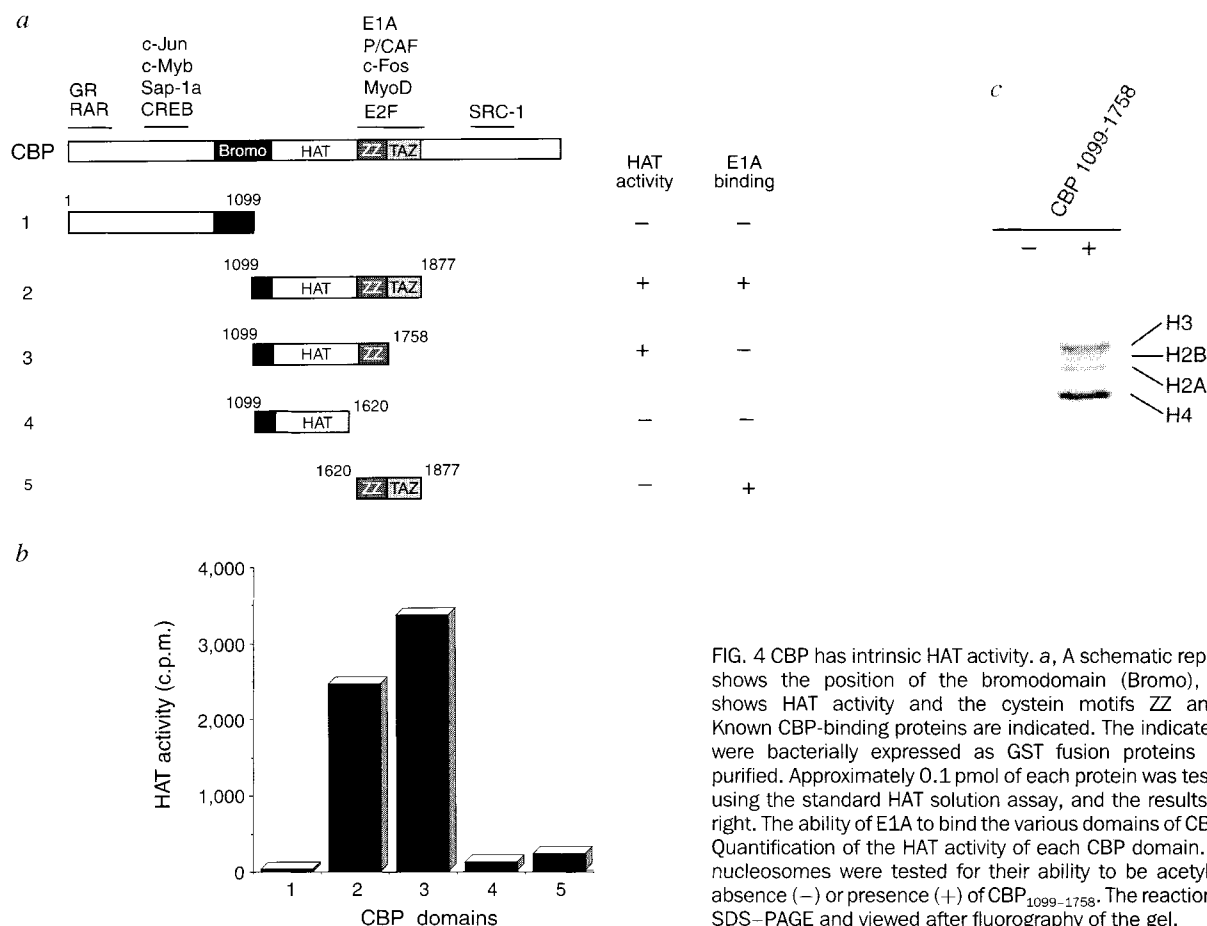


FIG. 4 CBP has intrinsic HAT activity. **a**, A schematic representation of CBP shows the position of the bromodomain (Bromo), the domain that shows HAT activity and the cysteine motifs ZZ and TAZ (ref. 11). Known CBP-binding proteins are indicated. The indicated domains of CBP were bacterially expressed as GST fusion proteins and subsequently purified. Approximately 0.1 pmol of each protein was tested for HAT activity using the standard HAT solution assay, and the results are shown on the right. The ability of E1A to bind the various domains of CBP is also shown. **b**, Quantification of the HAT activity of each CBP domain. **c**, Histones within nucleosomes were tested for their ability to be acetylated either in the absence (–) or presence (+) of CBP_{1099–1758}. The reactions were resolved by SDS–PAGE and viewed after fluorography of the gel.

ities: one intrinsic to CBP, and one associated with the P/CAF protein. E1A can displace P/CAF from CBP but does not mask the CBP-associated HAT activity. Thus the HAT functions that are intrinsic to the CBP protein might be pertinent to the ability of E1A to activate transcription and induce the S phase of the cell cycle. Some transcriptional activators bind CBP in a region that overlaps and possibly masks the P/CAF binding site (Fig. 4), which suggests that the intrinsic CBP HAT activity may be specifically relevant to these activators. □

Methods

HAT assays. Liquid HAT assays were performed using a modification of the assay previously described¹². Crude core histones (25 µg) and enzyme samples were mixed to give a final volume of 30 µl in buffer IPH (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% (by volume) NP-40, 0.1 mM PMSF). Reactions were initiated by the addition of [³H]-acetyl coA (3.7 kBq; 78.1 GBq mmol⁻¹; NEN) and the reactions incubated at 30 °C for 10–45 min. [³H]-acetyl incorporation into histones was determined by liquid scintillation counting of reactions spotted onto P-81 (Whatman) filters¹². For the analysis of GST-fusion HAT activity, ~0.1 pmol of each GST-fusion protein was incubated as described above. For the analysis of nucleosomes, ~5 µg of freshly prepared, H1-depleted nucleosomes (a gift from J. Thomas) were incubated as described above, except that 1 µl of [¹⁴C]-acetyl coA (1.85 kBq; 1.85 GBq mmol⁻¹; Amersham) was used. The reaction products were resolved by SDS–PAGE and viewed following fluorography of the gel.

IP–HAT assays. COS cells were grown in 14-cm culture dishes, collected by scraping off in 1 ml ice-cold PBS, and gently pelleted by centrifugation. The PBS was aspirated and the cell were resuspended into 1 ml lysis buffer IPH (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA, 0.5% (by volume) NP-40, 0.1 mM PMSF, aprotinin, leupeptin, pepstatin). The lysis mixture was incubated on ice for 20 min and cleared by centrifugation at 12,000g for 10 min at 4 °C. Relevant antibodies were added to 1 ml of extract and incubated at 4 °C for 2 h. Protein A-sepharose/Protein G-sepharose (a 50:50 mix, 15 µl total) was added and the mixture rotated slowly overnight at 4 °C. The immune complexes were pelleted by gentle centrifugation and washed three times with 1 ml lysis buffer IPH. After the final wash, the buffer was aspirated down to 30 µl. We added

1.25 µl of 20 mg ml⁻¹ histones and 1 µl of [³H]-acetylCoA, and a HAT assay was performed at 30 °C. Histone acetylation was measured using the P-81 filter assay described above.

GST fusion proteins. Various domains of CBP were cloned into the relevant pGex Vector (Pharmacia) using PCR or engineered restriction sites. Recombinant proteins were expressed in and purified from *Escherichia coli* as reported previously¹³.

In vivo expression plasmids. All Ad5 E1A12S constructs were expressed from plasmid pBJ919, which is an RSV-driven expression vector (a gift from H. Land). All E1A mutants have been described previously⁵.

Cell culture and transfections. COS cells were maintained in DMEM supplemented with 10% fetal calf serum and grown at 37 °C, 5% CO₂. The E1A-inducible cell line MT-3 and the parent line GM637 were also grown in the above media. MT-3 cell media also contained 400 µg ml⁻¹ G418. E1A expression was induced by the addition of CdCl₂ to a final concentration of 5 µM. Cell transfection was done using the calcium phosphate coprecipitation technique.

Received 20 November; accepted 26 November 1996.

1. Chakravarti, D. *et al.* *Nature* **383**, 99–103 (1996).
2. Chivria, J. C. *et al.* *Nature* **365**, 855–859 (1993).
3. Janknecht, R. & Nordheim, A. *Oncogene* **12**, 1961–1969 (1996).
4. Bannister, A. J., Oehler, T., Wilhelm, D., Angel, P. & Kouzarides, T. *Oncogene* **11**, 2509–2514 (1995).
5. Bannister, A. J. & Kouzarides, T. *EMBO J.* **14**, 4758–4762 (1995).
6. Yuan, W., Condorelli, G., Caruso, M., Felsani, A. & Giordano, A. *J. Biol. Chem.* **271**, 9009–9013 (1996).
7. Trouche, D., Cook, A. & Kouzarides, T. *Nucleic Acids Res.* **24**, 4139–4145 (1996).
8. Yang, X. J., Ogryzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. *Nature* **382**, 319–324 (1996).
9. Hagmeyer, B. M. *et al.* *Oncogene* **12**, 1025–1032 (1996).
10. Wong, H. K. & Ziff, E. B. *J. Virol.* **68**, 4910–4920 (1994).
11. Ponting, C. P., Blake, D. J., Davies, K. E., Kendrick-Jones, J. & Winder, S. J. *Trends Biol. Sci.* **21**, 11–13 (1996).
12. Brownell, J. E. & Allis, C. D. *Proc. Natl Acad. Sci. USA* **92**, 6364–6368 (1995).
13. Bannister, A. J., Cook, A. & Kouzarides, T. *Oncogene* **6**, 1243–1250 (1991).

ACKNOWLEDGEMENTS. We thank J. Thomas for nucleosomes; D. Lane, S. P. Jackson, M. Green and J. Pines for antibodies; and P. Angel for the E1A-inducible cell line.

CORRESPONDENCE and requests for materials should be addressed to T.K. (e-mail: tk106@ mole.bio.cam.ac.uk).

Structural basis for selective inhibition of cyclooxygenase-2 by anti-inflammatory agents

Ravi G. Kurumbail, Anna M. Stevens, James K. Gierse, Joseph J. McDonald, Roderick A. Stegeman, Jina Y. Pak, Daniel Gildehaus, Julie M. Miyashiro*, Thomas D. Penning*, Karen Seibert, Peter C. Isakson & William C. Stallings

G. D. Searle, 700 Chesterfield Parkway North, St Louis, Missouri 63198, USA

* G. D. Searle, 4901 Searle Parkway, Skokie, Illinois 63198, USA

PROSTAGLANDINS and glucocorticoids are potent mediators of inflammation. Non-steroidal anti-inflammatory drugs (NSAIDs) exert their effects by inhibition of prostaglandin production. The pharmacological target of NSAIDs is cyclooxygenase (COX, also known as PGH synthase), which catalyses the first committed step in arachidonic-acid metabolism^{1,2}. Two isoforms of the membrane protein COX are known³: COX-1, which is constitutively expressed in most tissues, is responsible for the physiological production of prostaglandins⁴; and COX-2, which is induced by cytokines, mitogens and endotoxins in inflammatory cells⁵, is responsible for the elevated production of prostaglandins during inflammation^{6,7}. The structure of ovine COX-1 complexed with several NSAIDs has been determined^{8–10}. Here we report the structures of unliganded murine COX-2 and complexes with flurbiprofen, indomethacin and SC-558, a selective COX-2 inhibitor, determined at 3.0 to 2.5 Å resolution. These structures explain the structural basis for the selective inhibition of COX-2, and demonstrate some of the conformational changes associated with time-dependent inhibition.

The enzymatic activity of COX involves bis-oxygenation of arachidonic acid to PGG₂, which is further reduced to PGH₂ in a peroxidase reaction by the same protein¹¹. NSAIDs act at the cyclooxygenase active site, and most inhibit both COX-1 and COX-2 with little specificity^{12,13}, leading to serious side effects such as gastric lesions and renal toxicity. COX-2-selective inhibitors have been identified that show potent anti-inflammatory activity *in vivo* with minimal gastric side effects^{6,14,15}. The murine COX-2 gene encodes for 607 amino acids, including 17 residues of the signal peptide. The current structure begins with Ala 18 (residue 33 in COX-1 numbering), the first residue of the mature protein (Fig. 1). For clarity, we will use the COX-1 numbering. The last residue resolved in the current model is residue 583; we were unable to locate the final 35 residues in the carboxy-terminal region as previously reported for COX-1 (ref. 8).

Amino-acid sequences of three COX-1 and four COX-2 enzymes are known. Comparison of any COX-1 with any COX-2 enzyme leads to a sequence identity of ~60%. The signal peptide of COX-1 is seven residues longer than that of COX-2, and the N terminus of COX-1 has an insertion of eight residues, whereas the C terminus of COX-2 has an 18-residue insertion. The only other difference is that COX-2 has a proline after Ile 106 at the junction of the membrane-anchoring helices C and D. This additional residue causes only minor perturbation in the local structure, and does not affect the overall topology of the enzyme. Consistent with the high sequence identity, the overall structures of COX-1 and COX-2 are highly conserved (r.m.s. deviation of 0.9 Å for all Cα atoms). The structure consists of three distinct domains: an N-terminal epidermal growth factor (EGF) domain, followed by a membrane-binding motif, and the C-terminal catalytic domain which contains the cyclooxygenase and peroxidase active sites. With 87% identity and strict sequence conservation in the

cyclooxygenase active site, the structure of human COX-2 is expected to be very similar to the murine enzyme.

Flurbiprofen (Fig. 2), a slow-binding competitive inhibitor of both COX-1 and COX-2, binds in the long hydrophobic channel and excludes substrate from the cyclooxygenase active site. The binding site of flurbiprofen in COX-2 is identical to that observed in COX-1 (refs 8, 10), and is constructed from helices D, 6, 8, 17 and the segment that follows helix 6. The carboxylate of the drug forms a salt bridge with the guanidinium group of Arg 120 and a hydrogen bond to Tyr 355. The distal aryl ring forms close van der Waals interactions with the main-chain atoms of Gly 526 and Ala 527, and stacks tightly against Tyr 385, which is active in the catalysis or suicide-inactivation of the enzyme^{16,17}. The distal phenyl ring of flurbiprofen also interacts with Ser 530, the residue that is selectively acetylated by aspirin^{4,18}. The fluorophenyl ring of the inhibitor makes fewer contacts with protein atoms in COX-2, and is involved in van der Waals interactions with Val 349 and the main chain of Ala 527. In COX-1, the fluorine interacts with the side chain of Ile 523, but no such interaction is observed in COX-2, which has a valine at the corresponding position. The α-methyl group of the *S*-stereoisomer is 3.4 Å away from Leu 359; in the case of the *R*-stereoisomer, there would be steric clash of the inhibitor with Tyr 355 (ref. 8).

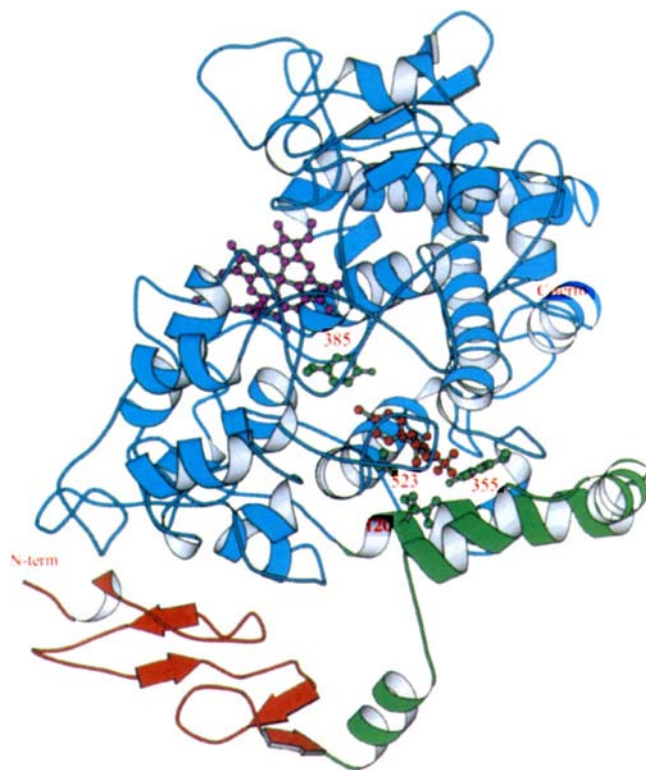


FIG. 1 Ribbon drawing of the structure of COX-2 complexed with the selective inhibitor SC-558. The EGF domain is red, the membrane domain green, and the catalytic domain blue. The N and C termini are labelled. Arg 120, Tyr 355, Tyr 385 and Val 523 are shown, as is the haem cofactor. The peroxidase active site is located close to the haem on the exterior surface of the protein. The inhibitor is bound in the cyclooxygenase active site. The secondary structural elements are assigned as follows: α helices, A, 74–80; B, 86–93; C, 97–104; D, 106–123; E, 139–143; 1, 174–182; 2, 196–209; 3, 238–244; 5, 296–319; 6, 325–353; 8, 379–384; 9, 411–417; 10, 420–427; 12, 445–457; 13, 463–469; 14, 478–482; 15, 485–495; 16, 503–508; 17, 520–535; 18, 553–560; 19, 564–570; 3₁₀ helices, a, 231–234; b, 266–269; c, 363–368; d, 387–390; e, 404–407; f, 538–541; g, 547–550, β-strands, 1-1, 46–48; 1-2, 56–58; 2-1, 63–65; 2-2, 71–73; 3-1, 255–257; 3-2, 260–262; 4-1, 395–397; 4-2, 400–402. The nomenclature for the helices is based on the structures of myeloperoxidase and COX-1 (ref. 8).

Indomethacin (Fig. 2), a classic non-selective cyclooxygenase inhibitor, binds deeply within the cyclooxygenase active site (Fig. 3b). Among the various complexes we examined, indomethacin penetrates furthest into the hydrophobic channel. Its chlorine atom interacts with Leu 384, which adopts a different side-chain conformation. The benzoyl group occupies an environment similar to that of the distal phenyl ring of flurbiprofen, and is stabilized by hydrophobic interactions with Leu 384, Phe 381, Tyr 385 and Trp 387. The benzoyl oxygen interacts with the side-chain hydroxyl of Ser 530 (3.6 Å) and with the Val 349 side chain. The benzoyl ring can adopt either *cis* or *trans* conformation with respect to the indole. The *cis* (Z) forms of benzylidenindene analogues of indomethacin are five times more potent than the *trans* (E) form¹⁹. In the complex with COX-2, indomethacin adopts a conformation similar to *cis* benzylidenindene. Preference for the *cis* isomer has been argued for, based on the COX-1 complex with indomethacin¹⁰. A 4-bromobenzyl analogue, which lacks the benzoyl oxygen, shows high selectivity for COX-2 compared to indomethacin²⁰, suggesting that the benzoyl oxygen is important in enhancing the affinity for COX-1. Whether this is achieved by direct contacts of the benzoyl oxygen with protein, or by conferring conformational rigidity to the inhibitor, is unclear. The carboxylate forms a salt bridge with Arg 120, and the indole ring interacts with Val 349 and Ser 353. Additional contacts are made with Tyr 355, Val 523 and Ala 527. The six-membered ring of indole interacts closely with main-chain atoms of Leu 352 and Ser 353, resulting in displacement of the 353–356 peptide segment by 0.75 Å because of a flexible glycine at 354. The *o*-methoxy group of indomethacin protrudes slightly into a relatively large cavity created in COX-2 adjacent to Ser 353, Tyr 355 and Val 523.

SC-558 (Fig. 2), a diaryl heterocyclic inhibitor with a 1,900-fold selectivity for COX-2 over COX-1, has a central pyrazole ring and a sulphonamide substituent attached to one of the aryl rings. Like flurbiprofen and indomethacin, SC-558 binds in the cyclooxygenase active site (Fig. 4a, b). The bromophenyl ring is bound in a hydrophobic cavity formed by Phe 381, Leu 384, Tyr 385, Trp 387, Phe 513 and Ser 530, with contributions from the backbone atoms of Gly 526 and Ala 527. The trifluoromethyl group is bound in an adjacent pocket formed by Met 113, Val 116, Val 349, Tyr 355, Leu 359 and Leu 531. These two features of SC-558 binding have rough equivalents in the binding of flurbiprofen and indomethacin to COX-2. The distal phenyl ring of flurbiprofen overlaps with the bromophenyl ring of SC-558; the pyrazole of SC-558 superim-

poses on the fluorophenyl ring of flurbiprofen. Unexpectedly, the carboxylate of flurbiprofen and the trifluoromethyl group of SC-558 bind in the same cavity (Fig. 5).

In COX-2, the channel that leads from membrane to the cyclooxygenase active site forks at the SC-558 binding site. One branch forms a cavity that accepts the bromophenyl ring of SC-558, whereas the other represents a cavity not observed in the COX-1 structure, and accommodates the entire phenylsulphonamide moiety. The phenyl ring is surrounded by hydrophobic residues Leu 352, Tyr 355, Phe 518, Val 523 and the backbone of Ser 353. Beyond this hydrophobic pocket, the sulphonamide group extends into a region near the surface of COX-2 that is relatively polar. The sulphonamide interacts with His 90, Gln 192 and Arg 513. The X-ray data do not allow unique definition of the conformation of the sulphonamide, but the conformer that is chosen is chemically reasonable. One of the oxygen atoms forms a hydrogen bond to His 90; the other oxygen is linked by a hydrogen bond to Arg 513. The amide nitrogen forms a hydrogen bond to the carbonyl oxygen of Phe 518.

The selectivity of SC-558 seems to result from the phenylsulphonamide moiety, which binds in a pocket that is more restricted in COX-1 and is unoccupied in complexes of COX-2 with non-selective inhibitors. This pocket branches off from the main channel that leads to the cyclooxygenase active site, and is more accessible in COX-2, primarily because of a substitution of isoleucine to valine at position 523. In COX-2, the smaller size of the valine side chain coupled with the conformational changes at Tyr 355 opens up the hydrophobic segment of the new pocket that comprises Leu 352, Ser 353, Tyr 355, Phe 518 and Val 523. The new pocket is also enlarged by the 0.7 Å movement of the Leu 352 to Tyr 355 peptide segment relative to that observed in the flurbiprofen complex. A similar pocket exists in COX-1, but is

TABLE 1 Crystallographic data and refinement

	Complex with				
	Flurbiprofen	Indomethacin	SC-558	SC-558	Unliganded
Space group	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>P</i> 2 ₁ 2 ₁ 2	<i>I</i> 222	<i>P</i> 2 ₁ 2 ₁ 2
Resolution limit (Å)	2.5	2.9	3.0	2.8	3.0
Observations	2,136,154	1,145,408	1,362,430	818,993	1,465,900
Unique reflections	76325	50058	36954	35324	52156
Completeness (%)	77.8	77.9	62.4	96.9	89.2
<i>R</i> _{sym} (%)	11.1	13.5	12.3	8.8	9.7
Refinement					
No. reflections with <i>F</i> > 2σ	59,815	41,254	30,300	31,066	41,397
Resolution (Å)	8–2.5	8–2.9	8–3.0	8–2.8	8–3.0
<i>R</i> _{cryst} (%)	23.6	21.9	21.5	22.0	23.1
<i>R</i> _{free} (%)	31.6	30.9	31.8	30.9	30.8
R.m.s. dev. bond lengths (Å)	0.014	0.014	0.014	0.014	0.015
R.m.s. dev. bond angles (deg)	1.8	1.8	1.9	1.8	1.9
Average <i>B</i> (Å ²)	15.9	12.5	9.9	20.0	12.3

* $R_{\text{sym}} = \sum |I_i - \langle I_i \rangle| / \sum I_i \times 100$, where I_i is the intensity of an individual measurement, and $\langle I_i \rangle$ is the mean intensity of that reflection.

FIG. 2 Chemical structures of the inhibitors: a, flurbiprofen; b, indomethacin; c, SC-558. Also shown are the IC₅₀ values of the inhibitors for human (h) COX-1 and COX-2, measured as discussed in ref. 26.

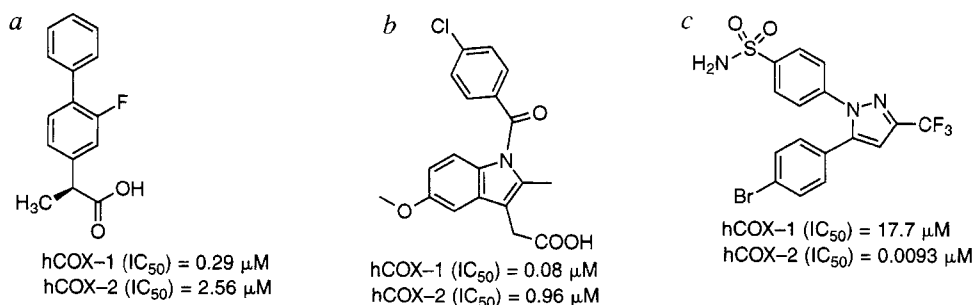
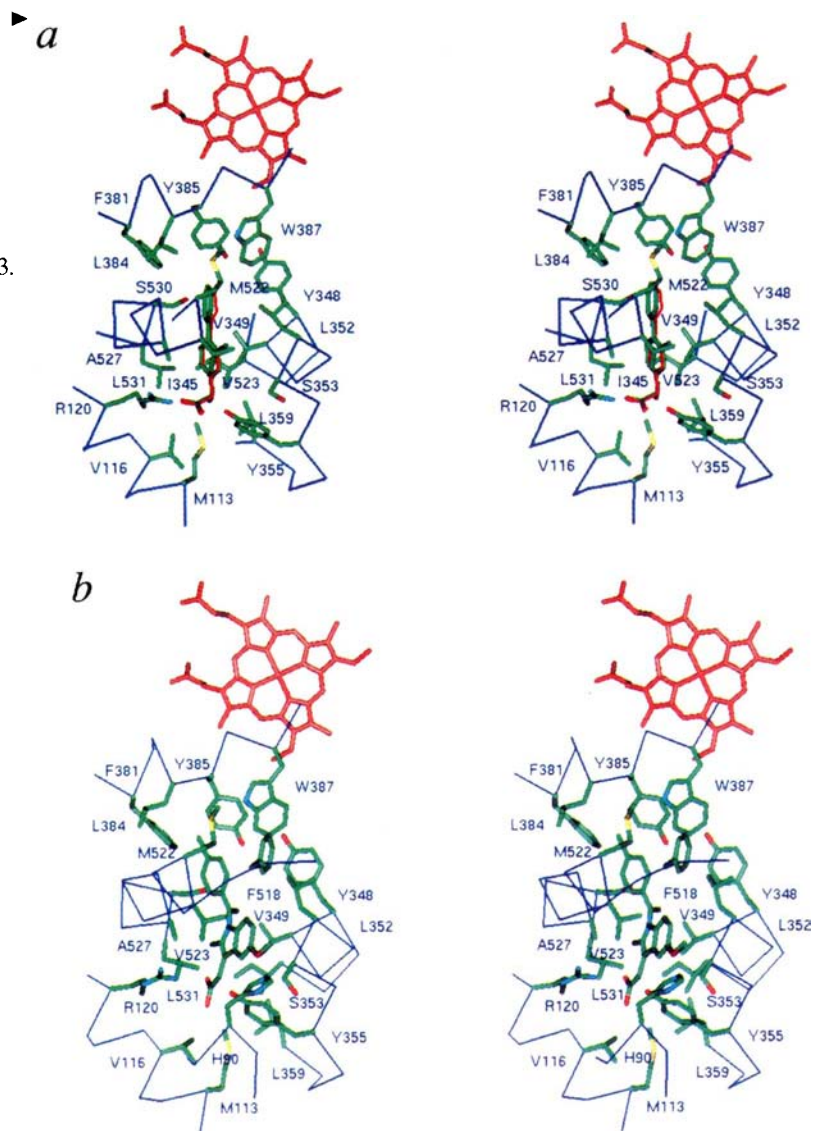


FIG. 3 *a*, Binding of flurbiprofen in the cyclooxygenase active site of COX-2. The stick figure (orange) was obtained by superimposing the C α atoms of the ovine COX-1 complex with flurbiprofen and displaying only the inhibitor atoms. *b*, Binding of indomethacin in the cyclooxygenase active site of COX-2. The methoxy group nearly extends into the 'selectivity pocket'.



inaccessible because of the bulkier isoleucine at 523, shown previously that the inhibition profile of human COX-2 is altered dramatically by mutation of a single amino acid²¹. COX-1 and the V523I variant of human COX-2 behave similarly on binding of selective inhibitors of the diaryl heterocyclic class. However, the potency of indomethacin for the variant remains unchanged, probably because indomethacin, a non-selective inhibitor, does not use the additional pocket present in COX-2.

Access of the phenylsulphonamide group to the new pocket in COX-2 is facilitated by another isoleucine to valine substitution. The side chain of the hydrophobic residue at 434 packs against Phe 518 which forms a molecular gate that extends to the new hydrophilic pocket; in COX-1, this gate is closed because of the larger isoleucine side chain. In COX-2, with the smaller side chain at 434, the gate has room to swing open, allowing the entry of the sulphonamide group. Thus the identity of the amino acid at position 434 also appears to make a significant contribution to selectivity in inhibitor binding. Near the surface of the protein, the inhibitor interacts with His 90, Gln 192 and Arg 513; His 90 and Gln 192 are conserved in the two isoforms. Arg 513 in COX-2 is replaced by histidine in COX-1; superposition of the two enzymes suggests that an imidazole ring at this position would not extend sufficiently for direct interactions with a sulphonamide group. Therefore the identity of the residue at position 513 is a third contributor to COX-2 specificity. Another residue in this pocket that might contribute to specificity is located at 516; it is alanine in COX-2, but serine in COX-1.

The primary determinant of selectivity of the diaryl heterocyclic class of inhibitors seems to be the phenylsulphonamide moiety. However, structure-activity relationships show that substitutions elsewhere can significantly alter the selectivity profiles for the two isoforms²². Similar results have been reported for indomethacin analogues²⁰. Thus a component of selectivity could be derived from structural differences between COX-1 and COX-2 outside the new pocket observed in COX-2. However, the differences we have observed around the inhibitor-binding sites of COX-1 and COX-2 are small, and found mainly in the position and conformations of helices B, D and 6 and the peptide segment that follows helix 6. The r.m.s. deviation for 44 C α atoms that surround flurbiprofen in the two structures, but do not form part of the new pocket, is only 0.37 Å. This leads us to conclude that diaryl heterocyclic compounds may have multiple modes of binding to cyclooxygenases.

Arg 120, the guanidinium group of which stabilizes the carboxylate of classical NSAIDs, is one of the few charged residues in the hydrophobic cyclooxygenase channel. However, in the structure with SC-558, which has no carboxylate group, there is no charge-charge interaction between the inhibitor and Arg 120. The lack of a carboxylate group in SC-558 could also be a significant component of its selectivity towards COX-2. Attempts to improve the potency against COX-2 by incorporating an acidic group on the pyrazole of the diaryl heterocyclic series results in poor selectivity (data not shown). The binding of SC-558 disrupts the salt bridge between Arg 120 and Glu 524. Because this interaction is absent,

Arg 120 displays higher *B*-factors, and there is a slight unwinding of the helix near Arg 120. Four of the helical hydrogen bonds are broken in the complex with the selective COX-2 inhibitor and helix D shifts by ~1.0 Å, closing up the mouth of the channel.

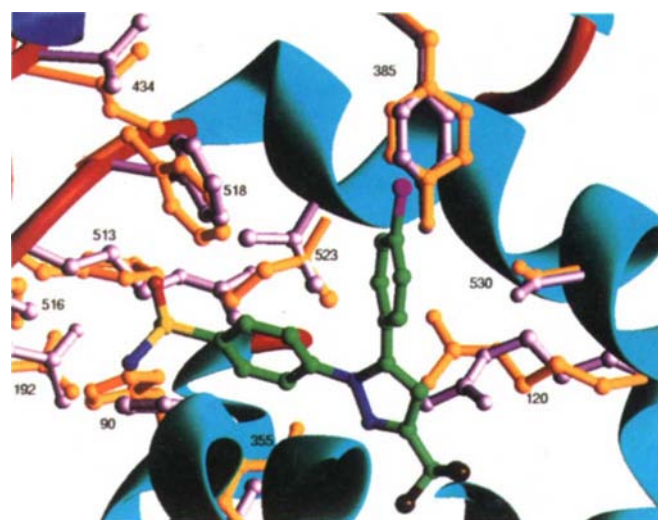
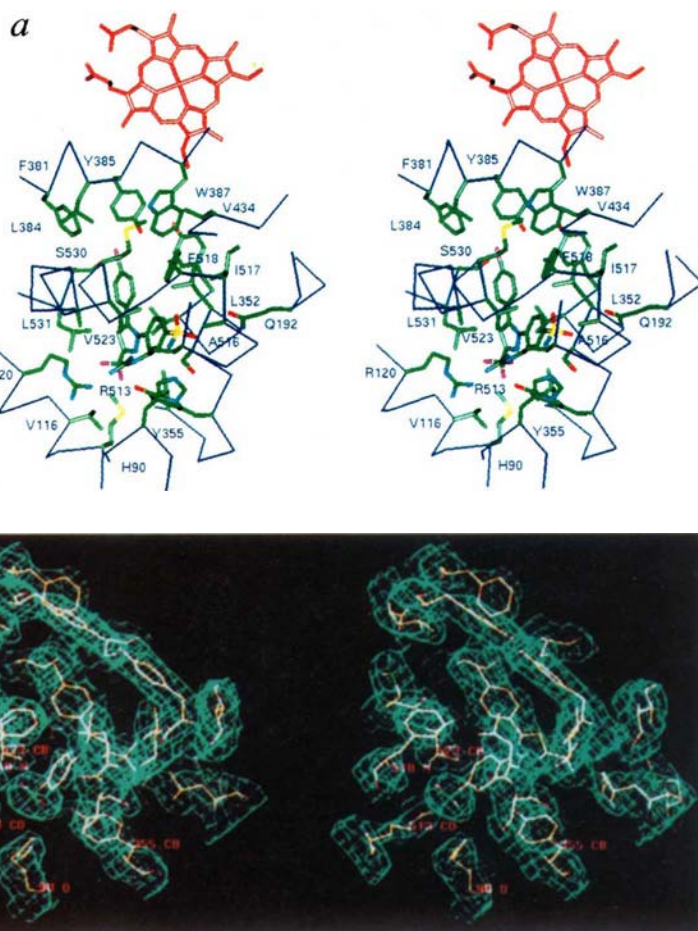
The unliganded enzyme has been the most difficult of these structures to obtain and is at the lowest resolution (3.0 Å). Kinetic measurements on slow-binding inhibitors such as indomethacin have led to the hypothesis that large-scale conformational changes must accompany the binding of the ligands²³. As a result, large differences between the free and the inhibited enzyme structures were anticipated. Overall, however, the structure of the free enzyme is very similar to those of the complexes with inhibitors (r.m.s. deviation of 0.4 Å for all C α atoms). The structural features that can account for the time-dependent inhibition must therefore be more subtle; dynamic, rather than static, conformational changes are probably involved in substrate entry and exit.

The binding of inhibitors does not greatly perturb the resting structure. Surprisingly, the channel mouth is more restricted (albeit only slightly) in the free enzyme. Arg 120 maintains its salt bridge with Glu 524, and changes in side-chain conformation are relatively minor. An exception is helix D, which forms part of the channel mouth; this helix is poorly ordered in the native structure, but not in the complexes. Evidence for the disorder in this helix is seen in the poorly defined features of electron-density maps and high thermal parameters (the average *B*-factor for

FIG. 4 *a*, Binding of SC-558 in the cyclooxygenase active site of COX-2. The phenyl sulphonamide ring extends into a new selectivity pocket that is more restricted in COX-1. The nitrogen atom of the pyrazole ring in SC-558 has a lone pair of electrons available for interactions, but does not form hydrogen bonds. *b*, Stereo view of the $2F_o - F_c$ electron-density map (contoured at 1σ) in the vicinity of SC-558 from the crystals with $I222$ symmetry. The map was calculated using reflections in the resolution range 20.0–2.8 Å with model phases. Superimposed on the map is the final model. Helix 17, which contains Val 523 and Ser 530, runs from top to bottom.

residues 118 to 122 is 29 Å^2 compared with 11.0 Å^2 for the entire structure). This is perhaps not surprising given the residues Met 113, Val 116 and Arg 120 form direct protein–ligand contacts in the complexes. Flexibility in this region may be a functional component of ligand binding.

We propose a working hypothesis for the time-dependent inhibition of COX. The known inhibitors of cyclooxygenases can be broadly classified into four types. Aspirin is the only known example of the first kind, which irreversibly inactivates both COX-1 and COX-2 by acetylating an active-site serine^{4,18}; this covalent modification interferes with the proper binding of arachidonic acid in the active site⁹. The second type consists of reversible, competitive inhibitors of both isoforms. These drugs, which include mefenamate and ibuprofen, compete with substrate for the cyclooxygenase active site. Flurbiprofen and indomethacin belong to the third class, which cause a slow, time-dependent inhibition of COX-1 and COX-2 (ref. 23). The time dependence of inhibition may result from the formation of a salt bridge between the carboxylate of the drug and Arg 120 from helix D²⁴. In the absence of substrate or acidic inhibitors, helix D displays considerable flexibility; upon binding of most acid NSAIDs, helix D becomes more ordered, presumably because a new salt bridge forms. This perhaps represents the conformational change that has been postulated to occur when time-dependent inhibitors bind^{23,25}. SC-558 is a representative of the fourth class, which selectively inhibit COX-2. These are weak competitive inhibitors of COX-1 but inhibit COX-2 in a slow, time-dependent process²¹. The time-dependent kinetics observed with COX-2 seems to be correlated with the molecular complexities that accompany the penetration of a portion of the inhibitor into a new pocket of the



inducible enzyme. In contrast, these compounds inhibit COX-1 in a simple competitive manner, probably because they are unable to access the restricted pocket. These conclusions are supported by experiments with the V523I mutant of COX-2 that show competitive inhibition with the diaryl heterocyclic compounds. The selectivity of this class for COX-2 also seems to be correlated with the kinetics of time-dependent inhibition.

To our knowledge this is the first example of a membrane protein being successfully studied as a target in structure-based drug design. The overall structures of COX-1 and COX-2 are highly conserved, the significant difference between the two being the presence of a much larger NSAID binding site in COX-2, primarily as a result of the substitution of a valine for an isoleucine in the cyclooxygenase active site. The deletion of a methylene group at this position (523) in COX-2 allows access to an additional pocket that leads directly to the solvent. The selective COX-2 inhibitor exploits the additional pocket for enhanced binding through the phenylsulphonamide moiety. Another feature that clearly differentiates the selective COX-2 inhibitor from the classical NSAIDs is the lack of a carboxylate group. Conformational changes observed among the various structures suggest that time-dependent inhibition of cyclooxygenases may occur by more than one mechanism. □

◀ FIG. 5 Superposition of COX-1 (yellow) and COX-2 (purple) around SC-558. SC-558 is shown bound to COX-2 (carbon, green; nitrogen, blue; sulphur, yellow; oxygen, red; fluorine, dark brown; bromine, magenta) superimposed on structurally homologous residues of COX-1 from its complex with flurbiprofen (omitted for clarity). The larger NSAID binding pocket in COX-2 is clearly visible. Access to this 'side pocket' is restricted in COX-1 because of the larger isoleucine at 523.

Methods

Protein expression and purification. Murine COX-2 cDNA fragments containing the coding sequence were cloned into baculovirus transfer vector pVL1393 and expressed in cultured Sf21 insect cells, as described for the human enzyme²⁶. The COX-2 enzyme was purified on an anion exchange column equilibrated with 20 mM Tris-HCl, 0.4% CHAPS, pH 8.1. The eluted enzyme was concentrated tenfold and further purified on a Sephacryl S-200 column equilibrated with 25 mM Tris-HCl, 150 mM NaCl, 0.3% β -octylglucopyranoside (β -OG). It was concentrated to 10 mg ml⁻¹ after reconstitution with hemin.

Crystallization. The enzyme, containing 1 mM inhibitor, was dialysed overnight at 4 °C against 20 mM sodium phosphate, 100 mM NaCl, 0.6% β -OG, and an inhibitor concentration that is five times its IC₅₀. Crystals were grown at 18 °C using the vapour diffusion method by mixing an equal volume of protein solution and a reservoir solution containing 20–34% monomethyl PEG 550, 10–240 mM MgCl₂, 50 mM EPPS, pH 8.0. The final concentration of β -OG in the drop was adjusted to 0.6%. Crystals grew as hexagonal and rectangular rods over three weeks.

Data collection. The crystals of the native enzyme and the complexes with flurbiprofen, indomethacin and SC-558 belong to the space group P2₁2₁2 (a = 180.0, b = 134.0, c = 120.0 Å) with two dimers in the asymmetric unit. The complex with SC-558 also crystallized in the space group I222 with similar lattice parameters, but with one dimer in the asymmetric unit. Crystals were flash frozen to -160 °C. Diffraction data (Table 1) were recorded using a 30-cm Mar Research image plate coupled to a Rigaku RU200 rotating anode X-ray generator fitted with mirror optics. The reflection intensities were integrated using the program DENZO and scaled using the program SCALEPACK²⁷.

Structure determination and refinement. The structure of the COX-2 complex with flurbiprofen was determined initially by molecular replacement methods using the MERLOT package²⁸. The structure of COX-1 dimer was used as the search model⁸. The two crystallographically independent dimers in the asymmetric unit have the same orientation, as indicated by a single sharp solution (5.0 σ) to the cross-rotation function. The translation function searches yielded two unambiguous solutions, with the same y and z but different x coordinates, corresponding to the positions of the two dimers. The vector relating the positions of the two dimers was determined in a subsequent intermolecular translation search (56 σ). The overall crystal packing of COX-2 closely resembles a pseudo-body-centred orthorhombic lattice (pseudo I222), as confirmed by a prominent peak in the Patterson function at x = 0.47, y = 0.5 and z = 0.5 (peak height, 50% of the origin peak). In the initial maps, electron density was clearly visible for both the haem and the inhibitor that were excluded from the model. The model was improved by successive rounds of rigid-body, restrained positional, B-factor and simulated annealing refinements using X-PLOR with non-crystallographic symmetry restraints²⁹. This was followed by manual rebuilding using the program O³⁰ and further refinements with X-PLOR. Flurbiprofen was included in the final rounds of refinement. The refined COX-2 structure, omitting the flurbiprofen, was used as the starting model for refinements of the native structure and the complex with indomethacin and SC-558. The structures were refined using X-PLOR as described above. The structure of the COX-2 complex with SC-558 that crystallized in the I222 space group was also determined by molecular replacement methods. The search model consisted of the structure of COX-2 complex with the same inhibitor in the P2₁2₁2 space group but without the inhibitor. Rotation and translation function searches using the program MERLOT yielded unambiguous solutions. The model was refined using the same protocols.

Received 3 October; accepted 20 November 1996.

- Smith, W. L. & Marnett, L. J. in *Metal Ions in Biological Systems* (eds Sigel, H. & Sigel, A.) 163–199 (Marcel Dekker, New York, 1994).
- Otto, J. C. & Smith, W. L. *J. Lipid Mediat. Cell Signal.* **12**, 139–156 (1995).
- Vane, J. R. et al. *Proc. Natl Acad. Sci. USA* **91**, 2046–2050 (1994).
- DeWitt, D. L. et al. *J. Biol. Chem.* **265**, 5192–5196 (1990).
- Herschman, H. R. *Biochim. Biophys. Acta* **1299**, 125–140 (1996).
- Seibert, K. et al. *Proc. Natl Acad. Sci. USA* **91**, 12013–12017 (1994).
- Masferrer, J. L. et al. *Proc. Natl Acad. Sci. USA* **91**, 3228–3232 (1994).
- Picot, D., Loll, P. J. & Garavito, R. M. *Nature* **367**, 243–249 (1994).
- Loll, P. J., Picot, D. & Garavito, R. M. *Nature Struct. Biol.* **2**, 637–643 (1995).
- Loll, P. J., Picot, D., Ekabo, O. & Garavito, R. M. *Biochemistry* **35**, 7330–7340 (1996).
- Marnett, L. J. & Maddipati, K. R. in *Peroxidases in Chemistry and Biology* Vol 1 (eds Everse, K. E. & Grisham, M. B.) 293–334 (CRC Press, Boca Raton, FL, 1991).
- Mitchell, J. A., Akarasereenont, P., Thiemermann, C., Flower, R. J. & Vane, J. R. *Proc. Natl Acad. Sci. USA* **90**, 11693–11697 (1993).
- Laneville, O. et al. *J. Pharmacol. Exp. Ther.* **271**, 927–934 (1994).
- Gans, K. R. et al. *J. Pharmacol. Exp. Ther.* **254**, 180–187 (1990).
- Futaki, N. et al. *Prostaglandins* **47**, 55–59 (1994).
- Smith, W. L., Eling, T. E., Kulmacz, R. J., Marnett, L. J. & Tsai, A. *Biochemistry* **31**, 3–7 (1992).
- Dietz, R., Nastainczyk, W. & Ruf, H. H. *Eur. J. Biochem.* **171**, 320–327 (1988).
- Van, J. R. *Nature New Biol.* **231**, 232–235 (1971).
- Hoogsteen, K. & Trenner, N. R. *J. Org. Chem.* **35**, 521 (1970).
- Black, W. C. et al. *Bioorg. Med. Chem. Lett.* **6**, 725–730 (1996).
- Gierse, J. K. et al. *J. Biol. Chem.* **271**, 15810–15814 (1996).
- Bertenshaw, S. R. et al. *Bioorg. Med. Chem. Lett.* **5**, 2919–2922 (1995).
- Kulmacz, R. J. & Lands, W. E. M. *J. Biol. Chem.* **260**, 12572–12578 (1985).
- Bhattacharyya, D. K., Lecomte, M., Rieke, C. J., Garavito, R. M. & Smith, W. L. *J. Biol. Chem.* **271**, 2179–2184 (1996).

- Copeland, R. A. et al. *Proc. Natl Acad. Sci. USA* **91**, 11202–11206 (1994).
- Gierse, J. K. et al. *Biochem. J.* **305**, 479–484 (1995).
- Otinowski, Z. & Minor, W. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, UK, 1993).
- Fitzgerald, P. M. D. *J. Appl. Crystallogr.* **21**, 273–278 (1988).
- Brunger, A. T. *XPLOR (Version 3.1): A System for X-ray Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1993).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* **A47**, 110–119 (1991).

ACKNOWLEDGEMENTS. We thank J. Edwards for growing crystals of COX-2; H.-S. Shieh for help with computations; C. Koboldt for measurement of IC₅₀; and R. M. Garavito for providing coordinates of ovine COX-1.

CORRESPONDENCE and requests for materials should be addressed to R.G.K. (e-mail: rgkuru@punchiri.monsanto.com).

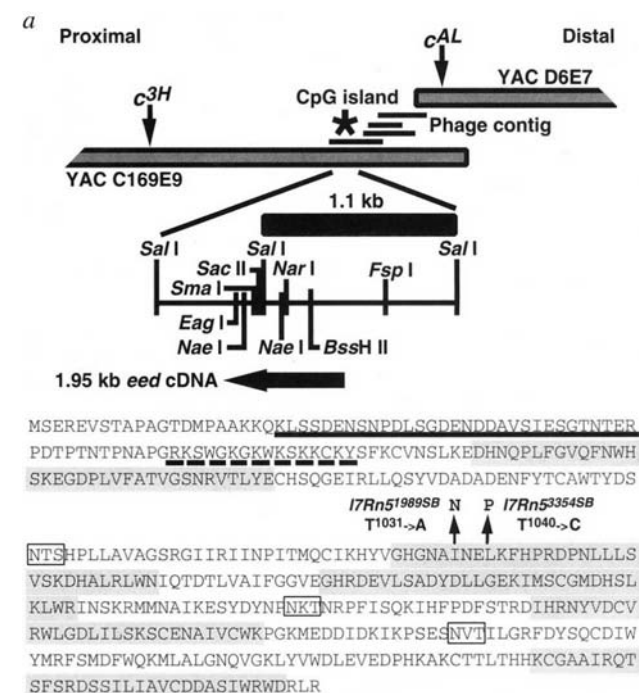
ERRATA

Positional cloning of a global regulator of anterior-posterior patterning in mice

Armin Schumacher, Cynthia Faust & Terry Magnuson

Nature **383**, 250–253 (1996)

THE amino-acid sequence was accidentally omitted from Fig. 1a. The complete figure is shown here. □

A transcriptional partner for MAD proteins in TGF- β signalling

Xin Chen, Melissa J. Rubock & Malcolm Whitman

Nature **383**, 691–696 (1996)

IN Fig. 1c of this Article, the shading of 1 H280 should have been striped as in the region of 1 H208 immediately above it, and not shaded grey as published. □